Chapter 2

Matrices

2.1 Linear Systems and Matrices

A typical *system of linear equations* is

$$\begin{aligned} 3x + 2y + z &= 5, \\ x \qquad - z &= 1, \\ x + \tfrac{1}{2}y + z &= 0. \end{aligned}$$

Let us consider it just a wee bit more abstractly, regarding the constant terms as parameters rather than definite numbers:

$$\begin{aligned} 3x + 2y + z &= a, \\ x \qquad - z &= b, \\ x + \tfrac{1}{2}y + z &= c. \end{aligned}$$

There are two things that one is accustomed to doing with such a list of formulas. Sometimes we *solve* them: given particular numbers as a, b, c, the task is to find x, y, z. On other occasions, however, we just *use* the formulas as they stand: given x, y, z, we calculate a, b, c. (That is, the formulas define a function from $\mathbf{R}^3$ into itself.) In what follows we shall be investigating both of these roles, first at the calculational and later at a more theoretical level.

To manipulate a linear system efficiently it is helpful to concentrate on the numerical coefficients in the left-hand sides of the equations. That is, we temporarily discard the variables x, y, z and the algebraic symbols $+$ and $=$, which actually carry very little information, and write just a table of numbers

$$A = \begin{pmatrix} 3 & 2 & 1 \\ 1 & 0 & -1 \\ 1 & \frac{1}{2} & 1 \end{pmatrix}.$$

A is called the *coefficient matrix* of the system. Similarly, we can write the numbers on the right-hand sides as a column (a vector in $\mathbf{R}^3$),

$$\begin{pmatrix} 5 \\ 1 \\ 0 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} a \\ b \\ c \end{pmatrix}. \tag{$*$}$$

It is helpful to introduce a more systematic notation: Instead of $(*)$ the column vector is written as

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \quad \text{or just} \quad \vec{y}.$$

The numbers in the matrix A are given names by

$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}.$$

Finally, the variables x, y, and z also form a vector,

$$\vec{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

The linear system thus is written

$$\begin{aligned} 3x_1 + 2x_2 + x_3 &= y_1\,, \\ x_1 \qquad\quad - x_3 &= y_2\,, \\ x_1 + \tfrac{1}{2}x_2 + x_3 &= y_3\,, \end{aligned}$$

or even, when we want to emphasize its abstract structure,

$$\begin{aligned} y_1 &= A_{11}x_1 + A_{12}x_2 + A_{13}x_3\,, \\ y_2 &= A_{21}x_1 + A_{22}x_2 + A_{23}x_3\,, \\ y_3 &= A_{31}x_1 + A_{32}x_2 + A_{33}x_3\,. \end{aligned}$$

We write all the equations at once as

$$y_i = A_{i1}x_1 + A_{i2}x_2 + A_{i3}x_3 \qquad (i = 1,\ 2,\ 3),$$

or, even more compactly,

$$y_i = \sum_{j=1}^{3} A_{ij} x_j \qquad (i = 1,\ 2,\ 3).$$

More generally, there could be n x's and m y's.

So, a *matrix* is a rectangular table of numbers. They can be any real numbers (or even complex), although in examples they'll often be integers for convenience. An $m \times n$ matrix has m rows and n columns.

$$\begin{pmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & & \\ \vdots & & \ddots & \vdots \\ A_{m1} & & \dots & A_{mn} \end{pmatrix} = (A_{ij}) = A.$$

The numbers are called *entries* or *elements* of the matrix. Note that the first index (i in the example) is the *row* index and the second is the *column* index. This general principle of notation, "ROWS BEFORE COLUMNS," was also used in describing the shape of the matrix as "$m \times n$".

Associated with such a matrix is a set of formulas for m dependent variables (the components of the vector variable $\vec{y}$) in terms of n independent variables (the components of $\vec{x}$):

$$y_i = \sum_{j=1}^{n} A_{ij} x_j \qquad (i = 1,\ 2, \dots,\ m).$$

If the ys have known values, this is a set of m equations in n unknowns. If the xs have known values, it is a formula defining a function from $\mathbf{R}^n$ into $\mathbf{R}^m$ (as we'll discuss in depth in Sec. 3.2). In this context one usually thinks of the variables as being arranged in *column vectors*, or matrices with just one column apiece:

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \qquad \vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}.$$

Then the entire system of linear equations is compressed to

$$\vec{y} = A\vec{x}.$$

One can think of a matrix as a kind of milling machine: You turn the input vector, $\vec{x}$, over on its side and drop it into the matrix; the matrix elements grind and mix the numbers in $\vec{x}$ and send the finished product, $\vec{y}$, out the m slots on the side. Less picturesquely, the simple expression $A\vec{x}$ is, by definition, an abbreviation for the list of the m sums $\sum_{j=1}^{n} A_{ij}x_j$. (This is the most elementary case of *matrix multiplication*, which is discussed in generality in Sec. 2.2. You can also think of $\vec{y} = A\vec{x}$ as a variant of the standard functional notation $\vec{y} = A(\vec{x})$.)

Solving linear equations

A linear system may have a *unique solution*:

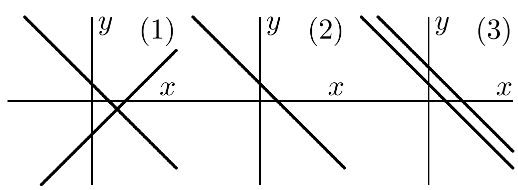
$$\begin{aligned} x + y &= 1, \\ x - y &= 2; \end{aligned} \tag{1}$$

or *many solutions*:

$$\begin{aligned} x + y &= 1, \\ 2x + 2y &= 2; \end{aligned} \tag{2}$$

or *no solutions*:

$$\begin{aligned} x + y &= 1, \\ x + y &= 2. \end{aligned} \tag{3}$$



That each system has the claimed property is fairly obvious for these two-variable examples, and becomes even clearer when one plots the graphs of the two equations in each case. The graph of each linear equation in each system is a line in the x–y plane. In case (1) the two lines intersect in a point (namely, $\left(\frac{3}{2}, -\frac{1}{2}\right)$), so there is exactly one choice of x and y that satisfies both equations simultaneously; this is the usual or "generic" situation. The two equations in (2) actually describe the same line; that is, the two lines in the problem coincide, and all the pairs (x, y) that are coordinates of points on that line are possible solutions of the system. In algebraic terms, the two equations are *not independent*. In (3) the two lines are parallel but distinct; they don't intersect at all. These two equations are *inconsistent*.

When there are more than two variables, one can't always tell just by looking whether a system has just one solution, or many, or none. For example, in dimension 3 the graph of each equation is a plane, so a system of three equations in three unknowns corresponds geometrically to three planes in $\mathbf{R}^3$. Three planes can intersect in a point, a line, a plane, or not at all. Furthermore, it is possible for three planes to have no intersection even though no two of the planes are parallel to each other.

Example 1. The system

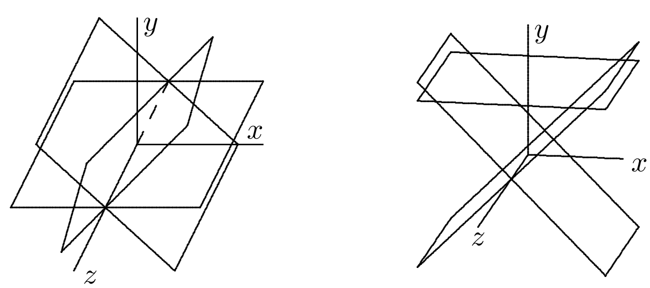
$$\begin{aligned} x + y &= 0, \\ x - y &= 0, \\ y &= 0 \end{aligned}$$

describes three planes in (x, y, z)-space that intersect in a line, the z axis. Each equation in the system fails to be independent of the other two (a fact that you can easily demonstrate algebraically).

Example 2. The three planes

$$\begin{aligned} x + y &= 0, \\ x - y &= 0, \\ y &= 4 \end{aligned}$$

have empty intersection. The intersection of each pair of planes is a line parallel to the z axis. The system of equations is inconsistent.



Example 1 Example 2

By rotating one of these sets of planes one would get an arrangement that is geometrically of exactly the same type, but is described by a more complicated set of three equations, hard to analyze without doing some calculations.

Example 3. The equations

$$\begin{aligned} 3x + 2y - 4z &= 1, \\ 6x + 4y - 8z &= 2, \\ -\frac{3}{5}x - \frac{2}{5}y + \frac{4}{5}z &= -\frac{1}{5} \end{aligned}$$

all represent the same plane. Therefore, the intersection is that plane, and any (x, y, z) on the plane is a solution of the system.

Our first main order of business is a technique for solving linear equations, variously called "the matrix method" or *row reduction* or *(Gaussian) elimination*. It is a streamlining of what in high school is called "the method of addition and subtraction". Recall that if you're given the system

$$\begin{aligned} x - y &= 1, \\ 2x + 3y &= 0, \end{aligned} \tag{4}$$

you can adjust the coefficients so that some of them cancel:

$$\begin{aligned} 2x - 2y &= 2, \\ \underline{2x + 3y} &\underline{= 0} \\ -5y &= 2. \end{aligned}$$

Thus $y = -\frac{2}{5}$. Then you can add this equation to the first of equations (4) to get $x = \frac{3}{5}$.

Let us *systematize this method* using matrices: Our example system (4) has the coefficient matrix

$$\begin{pmatrix} 1 & -1 \\ 2 & 3 \end{pmatrix}.$$

On its right, we tack on the column vector of the "right-hand sides" of the equations:

$$\left(\begin{array}{cc|c} 1 & -1 & 1 \\ 2 & 3 & 0 \end{array}\right).$$

This is called the *augmented matrix* for the system. On it we will perform so-called *elementary row operations*, which are equivalent to the things high-school students do to the equations in the process of solving them, except that we don't bother to write down all the variables, plus signs, and so on — just the numbers. The allowed operations are:

1. Multiply a row by a nonzero constant.
2. Add a multiple of one row to another row.
3. Interchange two rows. (This corresponds to just changing the order in which the equations are listed.)

Let's do this for the example:

$$\begin{pmatrix} 1 & -1 & 1 \\ 2 & 3 & 0 \end{pmatrix} \xrightarrow{(2) \leftarrow (2) - 2(1)} \begin{pmatrix} 1 & -1 & 1 \\ 0 & 5 & -2 \end{pmatrix} \xrightarrow{(2) \leftarrow \frac{1}{5}(2)}$$

$$\begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & -\frac{2}{5} \end{pmatrix} \xrightarrow{(1) \leftarrow (1) + (2)} \left(\begin{array}{cc|c} 1 & 0 & \frac{3}{5} \\ 0 & 1 & -\frac{2}{5} \end{array}\right)$$

The result means

$$x = \tfrac{3}{5}, \qquad y = -\tfrac{2}{5}.$$

The notation

$$\xrightarrow{(2) \leftarrow (2) - 2(1)}$$

is shorthand for "Subtract twice the first row from the second row, and replace the second row by the result." Such an indication of which row operation you're performing at each step is a great help to somebody trying to follow your work (and a good way of preserving partial credit on exams if your arithmetic is not infallible).

The point is that the elementary operations don't change the *solutions* of the system. They do replace the equations by *equivalent* equations whose solutions are obvious.

To decouple the variables in this fashion, we don't need to guess which operations to perform. There is a systematic procedure, called putting a matrix into *reduced* or *row echelon form*. We state it, and demonstrate it with a 3×3 example:

$$\left(\begin{array}{ccc|c} 0 & 1 & -1 & 3 \\ 1 & -1 & 1 & 0 \\ 2 & 2 & 2 & 1 \end{array}\right).$$

Gauss–Jordan Algorithm: These are the steps to follow in row-reducing a matrix.

1. If necessary, interchange two rows so that the first column containing any nonzero elements has a nonzero element at the top. (Additional interchanges of rows are permissible for arithmetic convenience. You'd like the row that you'll multiply and subtract from other rows to be full of 0's and 1's, if possible.)

$$\xrightarrow[(1)\leftrightarrow(2)]{} \begin{pmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & -1 & 3 \\ 2 & 2 & 2 & 1 \end{pmatrix}$$

2. Divide the first row by its first nonzero element (changing that to a 1). This nonzero element is called the *pivot* for the column in question. (Again, arithmetic convenience may suggest multiplying or dividing other rows by appropriate constants at any time.)

(Step 2 is unnecessary in the example.)

3. Clear out all other nonzero elements in that first column by subtracting from each row an appropriate multiple of the top row.

$$\xrightarrow[(3)\leftarrow(3)-2(1)]{} \begin{pmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & -1 & 3 \\ 0 & 4 & 0 & 1 \end{pmatrix}$$

4. Now do the same things to the submatrix below the top row and to the right of the zeros. (In the example,

$$\begin{pmatrix} 1 & -1 & 3 \\ 4 & 0 & 1 \end{pmatrix}$$

is this submatrix.) Also clear out the nonzero elements *above* the new leading 1.

$$\xrightarrow[\substack{(3)\leftarrow(3)-4(2)\\(1)\leftarrow(1)+(2)}]{} \begin{pmatrix} 1 & 0 & 0 & 3 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & 4 & -11 \end{pmatrix}$$

5. Continue in this manner until you reach either the bottom or the right edge of the matrix.

$$\xrightarrow[\substack{(3)\leftarrow\frac{1}{4}(3)\\(2)\leftarrow(2)+(3)_{\text{new}}}]{} \begin{pmatrix} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & \frac{1}{4} \\ 0 & 0 & 1 & -\frac{11}{4} \end{pmatrix}$$

Let's compare the original augmented matrix with its reduced form. This tells us that the system

$$\begin{aligned} y - z &= 3 \\ x - y + z &= 0 \\ 2x + 2y + 2z &= 1 \end{aligned}$$

is equivalent to the system

$$\begin{aligned} x \qquad\qquad &= 3 \\ y \qquad &= \tfrac{1}{4} \\ z &= -\tfrac{11}{4} \end{aligned}$$

— which is *the answer*.

If you have any doubt that they are equivalent, put the variables and arithmetic symbols back in and work through all the steps in the old-fashioned way, comparing with what we did here.

This process is prone to arithmetic errors, so it is important to *check your answers* at the end.

This system had a unique solution; but in other cases the reduced matrix will turn out to be something like this:

$$\left(\begin{array}{ccc|c} 1 & 0 & 3 & 5 \\ 0 & 1 & 4 & 7 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

The equivalent system is

$$\begin{aligned} x \qquad + 3z &= 5 \\ y + 4z &= 7 \\ 0 &= 0 \end{aligned}$$

There are not enough conditions here to determine a unique solution, so the original system of equations (whatever it was) has *many solutions*. It is easiest to solve the reduced system from the bottom up: The lowest nontrivial equation in the stack expresses y in terms of z, so we let z be an arbitrary parameter, then use the two nontrivial equations to express y and x in terms of it. (In other 3×3 problems, both y and z are arbitrary, and the one non-trivial equation expresses x in terms of them.) Thus the complete solution is

$$\begin{aligned} &x = 5 - 3z, \\ &y = 7 - 4z, \\ &z \text{ arbitrary.} \end{aligned}$$

(Some people consider it clearer to introduce a completely new variable — let's call it t — and write

$$\begin{aligned} &x = 5 - 3t, \\ &y = 7 - 4t, \\ &z = t. \end{aligned}$$

Use whichever style you prefer. The second style exhibits the solution as a parametrized line — see Sec. 1.2.)

Finally, suppose that the reduced matrix had been

$$\left(\begin{array}{ccc|c} 1 & 0 & 3 & 5 \\ 0 & 1 & 4 & 7 \\ 0 & 0 & 0 & 4 \end{array}\right)$$

(the same as the previous example except for the bottom right element). Then the bottom equation of the equivalent system would be $0 = 4$, which can't be satisfied. In such a case, the system we started with *has no solutions.*

Historical remark: Solution of linear systems by row reduction is demonstrated in a Chinese manuscript from roughly 200 B.C. (2000 years before Gauss and Jordan). Actually, the Chinese did *column* reduction, because they wrote the equations *down* the page like any other Chinese sentence.

More examples and variations

Example 4. Let us work through a 4×4 system. This is large enough that the advantages in efficiency of the Gauss–Jordan elimination method over other methods becomes quite clear. In following examples and working exercises of this magnitude, you may want to use a symbolic manipulation (computer algebra) program, such as *Mathematica* or *Maple*, to perform the row operations. (The libraries distributed with such programs usually also include commands that perform a complete row reduction, matrix inversion, or solution of a linear system at one fell swoop; these are, of course, quite useful for practical purposes, but we hope that you will not defeat the educational purpose of the exercises by using them now.) Our problem is to solve the system

$$\begin{aligned} x_1 + 2x_2 + 3x_3 + 4x_4 &= 7, \\ 3x_1 + 3x_2 + 5x_3 + 7x_4 &= 13, \\ 4x_1 + 4x_2 + 4x_3 + 6x_4 &= 14, \\ 3x_1 + \ x_2 - \ x_3 - 3x_4 &= 11. \end{aligned}$$

SOLUTION: Start reducing the augmented matrix:

$$\left(\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 7 \\ 3 & 3 & 5 & 7 & 13 \\ 4 & 4 & 4 & 6 & 14 \\ 3 & 1 & -1 & -3 & 11 \end{array}\right) \begin{array}{c} \longrightarrow \\ (2) \leftarrow (2) - 3(1) \\ (3) \leftarrow (3) - 4(1) \\ (4) \leftarrow (4) - 3(1) \end{array} \left(\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 7 \\ 0 & -3 & -4 & -5 & -8 \\ 0 & -4 & -8 & -10 & -14 \\ 0 & -5 & -10 & -15 & -10 \end{array}\right)$$

Now let's deviate from the strict Gauss–Jordan procedure for an "arithmetic convenience" step (getting rid of minus signs and some factors that are common to all the elements of a row):

$$\begin{array}{c} \longrightarrow \\ (2) \leftarrow -(2) \\ (3) \leftarrow -\frac{1}{2}(3) \\ (4) \leftarrow -\frac{1}{5}(4) \end{array} \left(\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 7 \\ 0 & 3 & 4 & 5 & 8 \\ 0 & 2 & 4 & 5 & 7 \\ 0 & 1 & 2 & 3 & 2 \end{array}\right)$$

At this point, the algorithm instructs us to divide the second row by 3 and to subtract appropriate multiples of that row from the others. In the present case that leads to very tedious arithmetic with fractions. The arithmetic in this example is much easier if we first subtract the third row from the second; this produces two new zeros and has the accidental advantage of making the first nonzero element of the second row automatically a 1. (Although this is a legal step — it replaces the system with an equivalent one — such deviations from the algorithm are dangerous in inexperienced hands. Always keep in mind the overriding strategy of *systematically filling up the lower left corner of the matrix with zeros*, else you may find yourself in an endless, aimless circle of row operations. Note also that a computer doesn't care about tedious noninteger arithmetic, but does care very much about proceedures that are not precisely defined; therefore, a computer program to row-reduce matrices will probably *not* involve any of the optional steps we have called "arithmetic conveniences". (See Example 8, however.))

$$\xrightarrow[(2) \leftarrow (2)-(3)]{} \left(\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 7 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 2 & 4 & 5 & 7 \\ 0 & 1 & 2 & 3 & 2 \end{array}\right) \xrightarrow[\substack{(3) \leftarrow (3)-2(2) \\ (4) \leftarrow (4)-(2) \\ (1) \leftarrow (1)-2(2)}]{} \left(\begin{array}{cccc|c} 1 & 0 & 3 & 4 & 5 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 4 & 5 & 5 \\ 0 & 0 & 2 & 3 & 1 \end{array}\right)$$

$$\xrightarrow[\substack{(3) \leftarrow \frac{1}{4}(3) \\ (4) \leftarrow (4)-2(3)_{\text{new}} \\ (1) \leftarrow (1)-3(3)_{\text{new}}}]{} \left(\begin{array}{cccc|c} 1 & 0 & 0 & \frac{1}{4} & \frac{5}{4} \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & \frac{5}{4} & \frac{5}{4} \\ 0 & 0 & 0 & \frac{1}{2} & -\frac{3}{2} \end{array}\right)$$

$$\xrightarrow[\substack{(4) \leftarrow 2(4) \\ (1) \leftarrow (1)-\frac{1}{4}(4)_{\text{new}} \\ (3) \leftarrow (3)-\frac{5}{4}(4)_{\text{new}}}]{} \left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 5 \\ 0 & 0 & 0 & 1 & -3 \end{array}\right).$$

Thus the original system is equivalent to the system

$$\begin{aligned} x_1 \qquad\qquad\qquad &= 2, \\ x_2 \qquad\qquad &= 1, \\ x_3 \qquad &= 5, \\ x_4 &= -3, \end{aligned}$$

which is the answer.

Example 5. In an application the entries in a matrix are likely to be functions of various parameters in the problem. Therefore, it is necessary to do calculations with algebraic expressions, not just numbers. The nature of the answer may depend on the numerical values of the parameters, so the general solution must be stated in terms of a list of cases. Our example system is

$$\begin{aligned} x + \alpha y &= 2, \\ \alpha x + 4y &= 0. \end{aligned}$$

SOLUTION: The augmented matrix is

$$\begin{pmatrix} 1 & \alpha & 2 \\ \alpha & 4 & 0 \end{pmatrix}.$$

Subtract α times the first row from the second row:

$$\begin{pmatrix} 1 & \alpha & 2 \\ 0 & 4-\alpha^2 & -2\alpha \end{pmatrix}.$$

Case 1: $\alpha \neq \pm 2$. Divide the second row by $4-\alpha^2$:

$$\begin{pmatrix} 1 & \alpha & 2 \\ 0 & 1 & \frac{2\alpha}{\alpha^2-4} \end{pmatrix}.$$

We can now see that

$$y = \frac{2\alpha}{\alpha^2-4}.$$

Therefore,

$$x = 2 - \alpha y = 2 - \frac{2\alpha^2}{\alpha^2-4} = \frac{-8}{\alpha^2-4}.$$

Case 2: $\alpha = \pm 2$. Then the bottom row of the matrix represents the equation $0 = \mp 4$; the system has no solutions.

You will have noticed that in Example 5 we did not carry the row-reduction procedure all the way to the end. We stopped as soon as the task of finding the solutions was reduced to trivial algebraic substitutions. The row-reduction algorithm as stated above is strictly called *Gauss–Jordan elimination* (when applied to the augmented matrix of a system of equations). What we did in Example 5 was to ignore, in Step 4, the instruction "Also clear out the nonzero elements *above* the new leading 1." This less complete reduction algorithm is called *Gauss elimination*. That is enough to produce a set of equations of the type

$$\begin{aligned} x_1 &= \ldots x_2 \ldots x_3 \ldots, \\ x_2 &= \ldots x_3 \ldots \\ &\ldots \end{aligned}$$

that can be solved from the bottom up. The full Gauss–Jordan procedure requires more operations on the matrix, but it requires fewer steps of "back-substitution" in the final solution of the equations (none at all in the case when the system has a unique solution). We shall do the next two examples by the incomplete Gauss method.

Example 6. Solve the system

$$
\begin{aligned}
2x_1 + 5x_2 + x_3 - 2x_4 &= 5,\\
x_1 + 3x_2 + 4x_3 - 6x_4 &= -2,\\
5x_1 - 2x_2 + 5x_3 + 3x_4 &= -6,\\
4x_1 + 9x_2 - 5x_3 + 6x_4 &= 17.
\end{aligned}
$$

Solution:

$$
\left(\begin{array}{cccc|c} 2 & 5 & 1 & -2 & 5\\ 1 & 3 & 4 & -6 & -2\\ 5 & -2 & 5 & 3 & -6\\ 4 & 9 & -5 & 6 & 17 \end{array}\right)
\underset{(1)\leftrightarrow(2)}{\longrightarrow}
\left(\begin{array}{cccc|c} 1 & 3 & 4 & -6 & -2\\ 2 & 5 & 1 & -2 & 5\\ 5 & -2 & 5 & 3 & -6\\ 4 & 9 & -5 & 6 & 17 \end{array}\right)
$$

$$
\underset{\substack{(2)\leftarrow(2)-2(1)\\(3)\leftarrow(3)-5(1)\\(4)\leftarrow(4)-4(1)}}{\longrightarrow}
\left(\begin{array}{cccc|c} 1 & 3 & 4 & -6 & -2\\ 0 & -1 & -7 & 10 & 9\\ 0 & -17 & -15 & 33 & 4\\ 0 & -3 & -21 & 30 & 25 \end{array}\right)
\underset{\substack{(2)\leftarrow-(2)\\(3)\leftarrow(3)+17(2)_{\text{new}}\\(4)\leftarrow(4)+3(2)_{\text{new}}}}{\longrightarrow}
$$

$$
\left(\begin{array}{cccc|c} 1 & 3 & 4 & -6 & -2\\ 0 & 1 & 7 & -10 & -9\\ 0 & 0 & 104 & -137 & -149\\ 0 & 0 & 0 & 0 & -2 \end{array}\right)
$$

The last equation from the reduced augmented matrix is $0 = -2$. Therefore, *the system has no solution.*

Example 7. Solve the system

$$
\begin{aligned}
2x_1 + 3x_2 + x_3 - 5x_4 &= -1,\\
2x_1 \phantom{{}-12x_2} + 2x_3 - 4x_4 &= -4,\\
2x_1 - 12x_2 + 8x_3 - 4x_4 &= -4,\\
x_1 - 8x_2 + 5x_3 - 2x_4 &= -2.
\end{aligned}
$$

Solution:

$$
\left(\begin{array}{cccc|c} 2 & 3 & 1 & -5 & -1\\ 2 & 0 & 2 & -4 & -4\\ 2 & -12 & 8 & -4 & -4\\ 1 & -8 & 5 & -2 & -2 \end{array}\right)
\underset{(1)\leftrightarrow(4)}{\longrightarrow}
\left(\begin{array}{cccc|c} 1 & -8 & 5 & -2 & -2\\ 2 & 0 & 2 & -4 & -4\\ 2 & -12 & 8 & -4 & -4\\ 2 & 3 & 1 & -5 & -1 \end{array}\right)
$$

$$
\underset{\substack{(2)\leftarrow(2)-2(1)\\(3)\leftarrow(3)-2(1)\\(4)\leftarrow(4)-2(1)}}{\longrightarrow}
\left(\begin{array}{cccc|c} 1 & -8 & 5 & -2 & -2\\ 0 & 16 & -8 & 0 & 0\\ 0 & 4 & -2 & 0 & 0\\ 0 & 19 & -9 & -1 & 3 \end{array}\right)
\underset{\substack{(2)\leftarrow\frac{1}{16}(2)\\(3)\leftarrow(3)-4(2)_{\text{new}}\\(2)\leftarrow(4)-19(2)_{\text{new}}}}{\longrightarrow}
$$

$$\left(\begin{array}{cccc|c} 1 & -8 & 5 & -2 & -2 \\ 0 & 1 & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & -1 & 3 \end{array}\right) \underset{\substack{(3) \leftrightarrow (4) \\ (3) \leftarrow 2(3)_{\text{new}}}}{\longrightarrow} \left(\begin{array}{cccc|c} 1 & -8 & 5 & -2 & -2 \\ 0 & 1 & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & -2 & 6 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

The system corresponding to this reduced form of the original augmented matrix is

$$\begin{aligned} x_1 - 8x_2 + 5x_3 - 2x_4 &= -2, \\ x_2 - \frac{1}{2}x_3 \qquad &= 0, \\ x_3 - 2x_4 &= 6. \end{aligned}$$

Working from the bottom up, we take x_4 as an arbitrary parameter and solve successively for the others:

$$x_3 = 6 + 2x_4\,, \quad x_2 = 3 + x_4\,, \quad x_1 = -8.$$

The answer can also be written

$$\vec{x} = \begin{pmatrix} -8 \\ 3+t \\ 6+2t \\ t \end{pmatrix},$$

where t is an arbitrary real number.

Example 8. If you were writing a computer program to solve linear systems, rather than solving systems by hand, how might your strategy change? First of all, in Step 1 you should ignore the remark about "arithmetic convenience". It takes the computer just as long to add or multiply by 0 or 1 as by 3.059275. This is not the end of the story, however.

Solve

$$\begin{aligned} 10^{-30}x + \qquad y &= 1, \\ \frac{32\pi}{5}\,x + \frac{\pi^4}{2}\,y &= \sqrt{\frac{17}{19}}\,. \end{aligned}$$

Solution: Surely no sane person, solving this system by hand, would choose the second row as the pivot row "for arithmetic convenience". Instead, you would multiply the first row by 10^{+30} and proceed without row interchange. (Moreover, after finding y you would probably find x by back substitution, rather than doing a complete Gauss–Jordan reduction.) The exact answer is

$$x = 10^{30}(1-y),$$

where

$$y = \frac{\sqrt{\frac{17}{19}} - \frac{32\pi}{5}\,10^{30}}{\frac{\pi^4}{2} - \frac{32\pi}{5}\,10^{30}}\,.$$

Notice that y is very close to 1, and therefore x is very close to 0 (or is it?).

But since the exact numbers in this problem are so disgusting, you would probably solve this problem with a calculator. Converting the coefficients to decimals, we get

$$\begin{aligned} 10^{-30}x + \qquad\quad y &= 1, \\ 20.1062x + 48.7045y &= .945905 \end{aligned}$$

Reducing the augmented matrix, we get after two steps

$$\begin{pmatrix} 1 & 10^{30} & 10^{30} \\ 0 & -.201062 \times 10^{32} & -.201062 \times 10^{32} \end{pmatrix},$$

which leads to

$$y = 1.00000, \qquad x = 0.00000.$$

Check this by substituting into the second of the two original equations:

$$\begin{aligned} 0 + \frac{\pi^4}{2}\,1 &= \sqrt{\frac{17}{19}}\,; \\ 48.7045 &= .945905. \end{aligned}$$

Something has gone wrong. (It's called "roundoff error".)

Let's try again, reversing the order of the rows before doing a Gauss–Jordan reduction:

$$\begin{pmatrix} 20.1062 & 48.7045 & .945905 \\ 10^{-30} & 1 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1.00000 & 0.00000 & -2.37532 \\ 0.00000 & 1.00000 & 1.00000 \end{pmatrix}.$$

Thus

$$y = 1.00000, \qquad x = -2.37532.$$

Check:

$$\begin{aligned} \frac{32\pi}{5}\,(-2.37532) + \frac{\pi^4}{2}\,1 &= \sqrt{\frac{17}{19}}\,; \\ .945841 &= .945905. \end{aligned}$$

Much better!

This example shows that in numerical work it is dangerous to divide by a pivot number that is significantly smaller than other numbers in the problem. Interchange of rows is still useful, but for an entirely different reason than in exact hand calculation.

Example 9. We conclude with another system involving a parameter,

$$\begin{aligned} -\lambda x_1 + x_2 + x_3 &= 1, \\ x_1 - \lambda x_2 + x_3 &= -\lambda, \\ x_1 + x_2 - \lambda x_3 &= \lambda^2. \end{aligned}$$

SOLUTION: Form and reduce the augmented matrix:

$$\left(\begin{array}{ccc|c} -\lambda & 1 & 1 & 1 \\ 1 & -\lambda & 1 & -\lambda \\ 1 & 1 & -\lambda & \lambda^2 \end{array}\right) \underset{(1)\leftrightarrow(3)}{\longrightarrow} \left(\begin{array}{ccc|c} 1 & 1 & -\lambda & \lambda^2 \\ 1 & -\lambda & 1 & -\lambda \\ -\lambda & 1 & 1 & 1 \end{array}\right)$$

$$\underset{\substack{(2)\leftarrow(2)-(1) \\ (3)\leftarrow(3)+\lambda(1)}}{\longrightarrow} \left(\begin{array}{ccc|c} 1 & 1 & -\lambda & \lambda^2 \\ 0 & -\lambda-1 & 1+\lambda & -\lambda-\lambda^2 \\ 0 & 1+\lambda & 1-\lambda^2 & 1+\lambda^3 \end{array}\right) \equiv A_1 .$$

At this point the general algorithm instructs us to divide the second row by $-\lambda-1$. However, this is not possible if that quantity is zero. Therefore, we will need to treat that case separately, later.

Case I: $1+\lambda \neq 0$ (i.e., $\lambda \neq -1$).

$$\left(\begin{array}{ccc|c} 1 & 1 & -\lambda & \lambda^2 \\ 0 & -\lambda-1 & 1+\lambda & -\lambda-\lambda^2 \\ 0 & 1+\lambda & 1-\lambda^2 & 1+\lambda^3 \end{array}\right)$$

$$\underset{\substack{(2)\leftarrow\frac{-1}{1+\lambda}(2) \\ (3)\leftarrow\frac{1}{1+\lambda}(3)}}{\longrightarrow} \left(\begin{array}{ccc|c} 1 & 1 & -\lambda & \lambda^2 \\ 0 & 1 & -1 & \lambda \\ 0 & 1 & 1-\lambda & 1-\lambda+\lambda^2 \end{array}\right)$$

$$\underset{(3)\leftarrow(3)-(2)}{\longrightarrow} \left(\begin{array}{ccc|c} 1 & 1 & -\lambda & \lambda^2 \\ 0 & 1 & -1 & \lambda \\ 0 & 0 & 2-\lambda & 1-2\lambda+\lambda^2 \end{array}\right) \equiv A_2 .$$

Subcase A: $2-\lambda \neq 0$ ($\lambda \neq 2$).

$$\left(\begin{array}{ccc|c} 1 & 1 & -\lambda & \lambda^2 \\ 0 & 1 & -1 & \lambda \\ 0 & 0 & 2-\lambda & 1-2\lambda+\lambda^2 \end{array}\right) \underset{(3)\leftarrow\frac{1}{2-\lambda}(3)}{\longrightarrow} \left(\begin{array}{ccc|c} 1 & 1 & -\lambda & \lambda^2 \\ 0 & 1 & -1 & \lambda \\ 0 & 0 & 1 & \frac{(1-\lambda)^2}{2-\lambda} \end{array}\right).$$

The corresponding system is

$$\begin{aligned} x_1 + x_2 - \lambda x_3 &= \lambda^2, \\ x_2 - x_3 &= \lambda, \\ x_3 &= \frac{(1-\lambda)^2}{2-\lambda} . \end{aligned}$$

So

$$x_2 = x_3 + \lambda = \frac{(1-\lambda)^2}{2-\lambda} + \lambda = \frac{1}{2-\lambda}\,,$$

$$\begin{aligned} x_1 &= \lambda^2 - x_2 + \lambda x_3 = \lambda^2 - \frac{1}{2-\lambda} + \lambda\frac{(1-\lambda)^2}{2-\lambda} \\ &= \frac{2\lambda^2 - \lambda^3 - 1 + \lambda - 2\lambda^2 + \lambda^3}{2-\lambda} = \frac{\lambda-1}{2-\lambda}\,. \end{aligned}$$

Subcase B: $2 - \lambda = 0$ ($\lambda = 2$). The reduced matrix A_2 becomes

$$\left(\begin{array}{ccc|c} 1 & 1 & -2 & 4 \\ 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 1 \end{array}\right).$$

There are no solutions.

Case II: $1 + \lambda = 0$ ($\lambda = -1$). The augmented matrix A_1 is

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

There is only one equation, $x_1 + x_2 + x_3 = 1$. (In fact, all three of the original equations were exactly this.) If we take $x_2 = s$ and $x_3 = t$, where s and t are arbitrary, then $x_1 = 1 - s - t$.

SUMMARY OF ANSWER:

$\lambda \neq -1, \quad \lambda \neq 2 \;\Rightarrow \quad x_1 = \dfrac{\lambda-1}{2-\lambda}, \quad x_2 = \dfrac{1}{2-\lambda}, \quad x_3 = \dfrac{(1-\lambda)^2}{2-\lambda}$;

$\lambda = 2 \;\Rightarrow$ no solutions;

$\lambda = -1 \;\Rightarrow \quad x_1 = 1 - s - t, \quad x_2 = s, \quad x_3 = t,$ for arbitrary real numbers s and t.

Exercises

2.1.1 Write these linear systems in matrix form. (That is, find a matrix M and a vector $\vec{v}$ so that the system is expressed by $M\vec{x} = \vec{v}$.)

(a)

$$\begin{aligned} 5x + 3y &= 0, \\ 2x - 7y &= 0. \end{aligned}$$

(b)

$$\begin{aligned} 3x + 4y \qquad\quad + 2w &= 3, \\ 4x + 7y - 2z \qquad\quad &= 2, \\ 6x \qquad + 5z + 4w &= 5, \\ 4x - \ \ y + \ \ z + 2w &= 1. \end{aligned}$$

(c) $$ax_1 + bx_2 + cx_3 = 1, \qquad a^2x_1 + b^2x_2 + c^2x_3 = 0,$$

where a, b, and c are given real numbers ("parameters"), and x_1, x_2, and x_3 are the unknowns.

2.1.2 Write out in elementary notation the system of equations (if any) whose matrix form is

(a) $\begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}$

(b) $\begin{pmatrix} a & b \\ -b & a \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$

(c) $\begin{pmatrix} 1 & 2 \\ -1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

2.1.3 Solve the system

$$\begin{aligned} x - 2y + 2z &= 0, \\ 4x + 2y - \ \ z &= 2, \\ -x + \ \ y + 3z &= -1 \end{aligned}$$

(a) by the elementary addition-and-subtraction method;

(b) by row reduction (getting the same answer, of course).

2.1.4 Find all solutions of these systems.

(a)

$$\begin{aligned} w + \ \ x - 2y + 3z &= 0, \\ w + 2x - 8y + 2z &= -2. \end{aligned}$$

(b)

$$\begin{aligned} 3x + 2y + \ \ z - \ \ w &= 4, \\ x + \ \ y + \ \ z + \ \ w &= 1. \end{aligned}$$

(c)

$$\begin{aligned} x_1 + 3x_2 - x_3 &= 1, \\ x_1 - x_2 + x_3 &= 0, \\ 3x_1 + x_2 + x_3 &= 1 \end{aligned}$$

2.1.5 Find all solutions of these systems.

(a)

$$\begin{aligned} x - 2y + 3z &= 0, \\ x - 8y + 7z &= -2, \\ x + y + z &= 1. \end{aligned}$$

(b) $x + 2y + 3z + 4w = 1, \qquad x + y + z + w = 0.$

(c) $x - 2y + 3z = 0, \qquad x + y + z = 1.$

(d) $x + 2y = 1, \qquad 2x - 3y = 0, \qquad x + 4y = 2.$

2.1.6

(a) Find all solutions of

$$\begin{aligned} x - 3y + 2z &= 0, \\ x + 5y + 4z &= 10, \\ x + y + 3z &= 5. \end{aligned}$$

(b) Find all numbers b such that

$$\begin{aligned} x - 3y + 2z &= 0, \\ x + 5y + 4z &= 10, \\ x + y + 3z &= b \end{aligned}$$

has no solutions.

2.1.7 Find all solutions of

$$\begin{pmatrix} 1 & 2 \\ -1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ s \end{pmatrix}.$$

Distinguish between different cases for the parameter s.

2.1.8 Solve the system $M\vec{x} = \vec{v}$,

$$M = \begin{pmatrix} 3 & -5 & 2 & 4 \\ 10 & -9 & 3 & 7 \\ 8 & 2 & -2 & -2 \end{pmatrix}, \qquad \vec{v} = \begin{pmatrix} 2 \\ 7 \\ 5 \end{pmatrix}.$$

2.1.9 Solve the system $M\vec{x} = \vec{v}$,

$$M = \begin{pmatrix} 7 & 7 & 0 & -1 \\ 2 & 5 & -3 & 4 \\ 3 & 4 & -1 & 1 \end{pmatrix}, \qquad \vec{v} = \begin{pmatrix} 4 \\ -4 \\ -2 \end{pmatrix}.$$

2.1.10 Solve the system $M\vec{x} = \vec{v}$,

$$M = \begin{pmatrix} 3 & 4 & 0 & 2 \\ 4 & 7 & -2 & 0 \\ 6 & 0 & 5 & 4 \\ 4 & -6 & 7 & 2 \end{pmatrix}, \qquad \vec{v} = \begin{pmatrix} 3 \\ 2 \\ p \\ 1 \end{pmatrix},$$

where p is a parameter.

2.1.11 Solve by row reduction, using *Maple* or similar software to do the arithmetic and algebra at each step:

$$\begin{aligned} 6x_1 - 3x_2 + 2x_3 + 3x_4 + \ \ 5x_5 &= 3, \\ 10x_1 - 5x_2 + 3x_3 + 5x_4 + \ \ 7x_5 &= 4, \\ 2x_1 - \ \ x_2 + 3x_3 + 7x_4 + 11x_5 &= 8, \\ 2x_1 - \ \ x_2 \qquad\quad - \ \ x_4 - \ \ x_5 &= -1. \end{aligned}$$

2.1.12 Let $\hat{e}_1$ and $\hat{e}_2$ be the first two of the four natural basis elements of $\mathbf{R}^4$ — for example,

$$\hat{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

(a) Calculate the vectors $B\hat{e}_1$ and $B\hat{e}_2$, if

$$B = \begin{pmatrix} 1 & 0 & 10 & 9 \\ 2 & 6 & -7 & 8 \\ 5 & 0 & 1 & 9 \\ 2 & 7 & 0 & -2 \end{pmatrix}.$$

(b) Give a verbal description of $B\hat{e}_j$ for any j (and any B and any $\mathbf{R}^n$).

2.1.13 The general solution of the differential equation $y'' + y = 0$ is $y = c_1 \cos t + c_2 \sin t$. Find c_1 and c_2 to satisfy these boundary conditions:

(a) $y(0) = 1$, $y'(0) = -1$

(b) $y(0) = 0$, $y(\pi/2) = 0$

(c) $y(0) = 0$, $y(\pi) = 0$

(d) $y(0) = 0$, $y(\pi) = 1$

(e) $y(\pi/4) = 1$, $y'(\pi/4) = -2$

(f) $y(\pi/6) = -2$, $y(\pi/3) = 2$

2.1.14 The expression $y = c_1 e^{-t} + c_2 e^{-2t} + c_3 e^{-3t}$ is the general solution of some third-order homogeneous linear differential equation.

(a) Find the coefficients to satisfy $y(0) = 1$, $y'(0) = 2$, $y''(0) = -1$.

(b) Find a differential equation that has this expression as general solution. HINT: The roots of $(r+1)(r+2)(r+3) = 0$ are $r = -1$, -2, -3.

2.1.15 Let $\vec{u}_1 = (1, 2, 3)$ and $\vec{u}_2 = (3, 2, 1)$. *Suggestion:* In setting up systems of equations, think of all vectors in this exercise as *column* vectors, although they were not typeset that way.

(a) Express $\vec{w} = (4, 4, 4)$ as a linear combination of $\vec{u}_1$ and $\vec{u}_2$.

(b) Find an example of a vector in $\mathbf{R}^3$ that is *not* equal to a linear combination of $\vec{u}_1$ and $\vec{u}_2$.

2.2 Matrix Algebra

One of the advantages of representing lists of linear formulas by matrices is that we can perform various mathematical operations on a matrix, thought of as a single object.

Addition of two matrices is defined element-by-element:

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ -1 & 3 \end{pmatrix}.$$

The matrices must be of the same "shape" (both $m \times n$ with the same m and n). In formal generality,

$$(A+B)_{jk} \equiv A_{jk} + B_{jk} \quad \text{for all } j \text{ and } k.$$

This operation has an interpretation in terms of linear equations: Given two linear systems of the same shape and involving the same independent variables, we might have occasion to add the corresponding equations of the two systems and simplify the result by the distributive law (i.e., combine terms). For example, let

$$\begin{aligned} y_1 &= 2x_1 \\ y_2 &= \qquad\quad 2x_2\,, \end{aligned} \qquad (\vec{y} = A\vec{x}),$$

and

$$\begin{aligned} z_1 &= x_1 + x_2\,, \\ z_2 &= -x_1 + x_2\,, \end{aligned} \qquad (\vec{z} = B\vec{x});$$

then

$$\begin{aligned} y_1 + z_1 &= 2x_1 + (x_1 + x_2) = 3x_1 + x_2\,, \\ y_2 + z_2 &= 2x_2 + (-x_1 + x_2) = -x_1 + 3x_2\,; \end{aligned}$$

in matrix notation,

$$\vec{y} + \vec{z} = (A+B)\vec{x} \equiv A\vec{x} + B\vec{x}.$$

In other words, *addition of matrices corresponds to addition of the linear functions or formulas that the matrices represent.*

Multiplication by a number is also element-by-element:

$$3\begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 3 \\ 0 & 3 \end{pmatrix}; \quad (rA)_{jk} \equiv r\,A_{jk} \quad \text{for all } j \text{ and } k.$$

The usual laws of algebra apply:

Algebraic Identities 1:

$$\begin{aligned} A + B &= B + A, & r(A+B) &= rA + rB, \\ (A+B)+C &= A + (B+C), & (r+s)A &= rA + sA, \\ 1A &= A, & r(sA) &= (rs)A. \end{aligned}$$

The proofs of these laws are trivial: just apply the ordinary commutative, associative, and distributive laws for numbers to the individual elements of the matrices. (We give some numerical examples at the end of the section.) Incidentally, a list of properties like this should not be thought of as a thing to be memorized, but rather as something which one learns with practice to *use* automatically. Perhaps the most important things to learn consciously are the items that are *not* in the list; for example, in a moment we will see that *multiplication of two matrices is not commutative.*

The *zero matrix* satisfies $A + 0 = A$ for all A. In the 2×3 case, for example,

$$0 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The *negative* of a matrix is defined by $(-A)_{jk} = -A_{jk}$ (i.e., take the negative of each element) or by its basic property, $A + (-A) = 0$. *Subtraction* can be defined by

$$A - B \equiv A + (-B),$$

or it can be defined element-by-element.

Now for the really interesting part: *Matrix multiplication* is *not* element-by-element.

For two matrices to have a product, the number of columns of the first (left) matrix must equal the number of rows of the second one. Let's call this number p. Then the element of the product AB in the ith row and jth column is

$$(AB)_{ij} \equiv \sum_{k=1}^{p} A_{ik} B_{kj}\,.$$

For example,

$$\begin{pmatrix} 3 & 2 & -1 \\ 4 & 6 & 9 \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 4 & 4 \\ 0 & 43 & 19 \end{pmatrix}.$$

Note that the general pattern is

$$\begin{pmatrix} \xrightarrow{\hspace{6em}} \\ \xrightarrow{\hspace{6em}} \end{pmatrix} \begin{pmatrix} \big\downarrow & \big\downarrow \end{pmatrix}.$$

(This you *should* memorize immediately, or at least your fingers should.) Note also that $(AB)_{ij}$ is the dot product of the ith row of A with the jth column of B.

The product of an $m \times p$ matrix by a $p \times n$ matrix is an $m \times n$ matrix.

Note that $AB \neq BA$ in general. (They may not both be defined, because the shapes don't match up right. If they are both defined, the results may be of different sizes — $m \times n$ times $n \times m$ yields $m \times m$, but the opposite order yields $n \times n$. Finally, all of the matrices may be square and of the same size, but nevertheless the two products may be unequal; see examples in the exercises.)

This strange definition is not the product of someone's twisted imagination. It has fundamental interpretations in terms of linear equations:

1. *Substitution of one system into another:* If

$$\begin{aligned} y_1 &= x_1 + 2x_2\,, \qquad & x_1 &= -z_1 + 2z_2\,, \\ y_2 &= 3x_1 + x_2\,, \qquad & x_2 &= 3z_1 + z_2\,, \end{aligned}$$

then

$$\begin{aligned} y_1 &= (-z_1 + 2z_2) + 2(3z_1 + z_2) = 5z_1 + 4z_2\,, \\ y_2 &= 3(-z_1 + 2z_2) + (3z_1 + z_2) = 0 + 7z_2\,. \end{aligned}$$

In matrix notation:

$$\vec{y} = A\vec{x}, \qquad A \equiv \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix};$$

$$\vec{x} = B\vec{z}, \qquad B \equiv \begin{pmatrix} -1 & 2 \\ 3 & 1 \end{pmatrix};$$

thus

$$\vec{y} = C\vec{z}, \qquad C = \begin{pmatrix} 5 & 4 \\ 0 & 7 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} -1 & 2 \\ 3 & 1 \end{pmatrix} = AB.$$

We see that the arithmetic done in the substitution is precisely that involved in the definition of the matrix product.

In the formula $\vec{y} = AB\vec{z}$, the matrices act from right to left on the z_js. Note that $\vec{y}$ is a *composite function* of $\vec{z}$, like those encountered in elementary calculus:

$$h(z) = f(g(z)) \equiv f \circ g\,(z) = [\text{for example}]\ \sqrt{z+2}.$$

Multiplication of matrices corresponds to composition of the vectorial linear functions that the matrices represent.

2. *Application of a matrix to a vector:* When we write the system of linear formulas $y_j = \sum_k A_{jk}x_k$ as a single vectorial linear function, $\vec{y} = A\vec{x}$, we omit the parentheses that ordinarily surround a function's argument (independent variable). The historical reason for this notation is that $A\vec{x}$ can be interpreted as the matrix product of A with $\vec{x}$, the latter being regarded as a matrix with only one column:

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

From a "function" point of view, one would write $A\vec{x}$ as $A(\vec{x})$, and $\vec{y} = AB\vec{z}$ as $\vec{y} = A(B(\vec{x})) = (A \circ B)(\vec{x})$. The interpretation of matrices as linear functions will be treated in greater depth in Chapters 3 and 8.

Although it is not commutative, matrix multiplication does satisfy associative and distributive laws:

Algebraic Identities 2:

$$(AB)C = A(BC) \equiv ABC, \qquad (rA)B = A(rB) = r(AB) \quad \text{for numbers } r,$$

$$A(B+C) = AB + AC, \qquad (D+E)C = DC + EC,$$

whenever the matrices are of the right shapes for all the terms to make sense.

An *identity* (*unit*) *matrix*,

$$I = \begin{pmatrix} 1 & 0 & 0 & \dots \\ 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

represents the "trivial" linear system, $y_j = x_j$ for all j. (That is, $I\vec{x} = \vec{x}$.) It must be *square* ($n \times n$). Often people write "1" for I when there is no danger of confusion with the number 1. This convention is nice because it allows numerical multiples of the identity matrix to be represented simply by numbers:

$$5I \equiv 5 = \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix}$$

when it is understood that we are dealing with 2×2 matrices. (Some authors use special fonts, such as boldface or "blackboard bold", to distinguish the matrices 0 and 1 from the corresponding numbers.)

Algebraic Identities 3: For matrices of the appropriate shapes,

$$\begin{aligned} A+0&=A, & AI&=A,\\ A+(-A)&=0, & IA&=A. \end{aligned}$$

Powers of a square matrix are defined in the obvious way: $A^3 = AAA$, for instance. A numerical example is

$$\begin{pmatrix}1&2\\2&1\end{pmatrix}^3 = \begin{pmatrix}1&2\\2&1\end{pmatrix}\begin{pmatrix}5&4\\4&5\end{pmatrix} = \begin{pmatrix}13&14\\14&13\end{pmatrix}.$$

Furthermore, one defines $A^0 = I = 1$. Consequently, every *polynomial function* of a square matrix is defined; continuing the numerical example, we have

$$A^3-3A+5 = \begin{pmatrix}13&14\\14&13\end{pmatrix} - \begin{pmatrix}3&6\\6&3\end{pmatrix} + \begin{pmatrix}5&0\\0&5\end{pmatrix} = \begin{pmatrix}15&8\\8&15\end{pmatrix}.$$

In accordance with the discussion above, the resulting matrix represents a certain linear combination of substitutions of a system of linear formulas into itself — rather a mess if you try to write it out, but quite simple if you think of the system of linear formulas as a vectorial function, $\vec{y} = A\vec{x}$.

A is *diagonal* if $A_{jk} = 0$ whenever $j \neq k$. Example:

$$\begin{pmatrix}1&0&0\\0&1&0\\0&0&3\end{pmatrix}.$$

It represents a system in which the equations are *decoupled*:

$$y_1 = x_1\,, \qquad y_2 = x_2\,, \qquad y_3 = 3x_3\,.$$

Finally, the *transpose* of a matrix is defined by $(A^{\mathrm{t}})_{jk} \equiv A_{kj}\,.$

$$\begin{pmatrix}0&1\\2&3\end{pmatrix}^{\mathrm{t}} = \begin{pmatrix}0&2\\1&3\end{pmatrix}, \qquad \begin{pmatrix}1\\2\end{pmatrix}^{\mathrm{t}} = (\,1\quad 2\,) \equiv (1,2).$$

Unfortunately, the notation for transposes is not standardized. Other notations used for A^{t} include ${}^{\mathrm{t}}A$, A^{T}, $\tilde{A}$, A^*, $A^\dagger$. (The last two of these indicate a complex conjugation along with the transposition if the matrix elements are complex.)

The transpose operation satisfies its own list of identities, which we'll leave for Exercise 2.2.22. The most subtle of these is

$$(AB)^{\mathrm{t}} = B^{\mathrm{t}} A^{\mathrm{t}}.$$

A matrix is called *symmetric* if $A^{\mathrm{t}} = A$, *antisymmetric* or *skew-symmetric* if $A^{\mathrm{t}} = -A$. Examples:

$$\begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix} \quad \text{is symmetric,} \quad \begin{pmatrix} 0 & -2 \\ 2 & 0 \end{pmatrix} \quad \text{is antisymmetric.}$$

The following theorem is obvious once it has been pointed out:

Theorem: Every square matrix is the sum of a symmetric matrix and an antisymmetric matrix. Namely, $\frac{1}{2}(A+A^{\mathrm{t}})$ is always symmetric and $\frac{1}{2}(A-A^{\mathrm{t}})$ is always antisymmetric, and these two parts add up to A.

Economic applications

Example A. To produce a ton of steel requires 4 tons of coal and 2 tons of iron ore. To produce a ton of aluminum requires 10 tons of coal and 2 tons of bauxite. To produce a car requires a ton of steel and 1/4 ton of aluminum. Explain how these raw-material input requirements can be organized into matrices, and how the matrices could be used to calculate the amount of coal, iron ore, and bauxite needed to produce a certain number of cars.

Solution: The *input-output table* for the metal industry is a 3×2 matrix:

$$\begin{array}{l} \\ \text{coal} \\ \text{iron ore} \\ \text{bauxite} \end{array} \begin{array}{c} \begin{array}{cc} \text{steel} & \text{alum.} \end{array} \\ \begin{pmatrix} 4 & 10 \\ 2 & 0 \\ 0 & 2 \end{pmatrix} \end{array}.$$

This matrix represents the linear formulas which tell how much coal, ore, and bauxite are used in producing given amounts of steel and aluminum. (Note that the *input* for this *calculation* is the *output* of the *industrial process*, and vice versa.) Similarly, for the automotive industry we have the input-output matrix

$$\begin{array}{c c} & \text{cars} \\ \begin{array}{l} \text{steel} \\ \text{aluminum} \end{array} & \begin{pmatrix} 1 \\ .25 \end{pmatrix} \end{array}.$$

The matrix describing the composite production process, from natural resources to cars, is the product of these:

$$\begin{array}{c c} & \text{cars} \\ \begin{array}{l} \text{coal} \\ \text{iron ore} \\ \text{bauxite} \end{array} & \begin{pmatrix} a \\ b \\ c \end{pmatrix} \end{array} = \begin{pmatrix} 4 & 10 \\ 2 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ .25 \end{pmatrix}.$$

Example B. A car contains 2000 cubic inches of steel and 10 cubic inches of rubber. A bicycle contains 25 cubic inches of steel and 1 cubic inch of rubber. Steel weighs 2 pounds per cubic inch and costs \$3 per cubic inch. Rubber weighs 0.1 pound per cubic inch and costs \$4 per cubic inch. Organize these facts into matrices, and find the matrix that should be used to calculate the total weight and total cost of the material needed to make x cars and y bicycles.

Solution: Make the obvious abbreviations s, r, w, c. Then

$$\begin{pmatrix} s \\ r \end{pmatrix} = A \begin{pmatrix} x \\ y \end{pmatrix}, \quad \text{where } A = \begin{pmatrix} 2000 & 25 \\ 10 & 1 \end{pmatrix}.$$

(The top row expresses the fact that the total necessary steel is 2000 units for each car and 25 units for each bicycle. The second row says that the needed rubber is 10 units for each car and 1 unit for each bicycle.) Similarly,

$$\begin{pmatrix} w \\ c \end{pmatrix} = B \begin{pmatrix} s \\ r \end{pmatrix}, \quad \text{where } B = \begin{pmatrix} 2 & 0.1 \\ 3 & 4 \end{pmatrix}.$$

Therefore,

$$\begin{pmatrix} w \\ c \end{pmatrix} = BA \begin{pmatrix} x \\ y \end{pmatrix},$$

where

$$BA = \begin{pmatrix} 2 & 0.1 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 2000 & 25 \\ 10 & 1 \end{pmatrix} = \begin{pmatrix} 4001 & 50.1 \\ 6040 & 79 \end{pmatrix}.$$

Additional examples

Example 1. *Commutative law of addition.*

$$A = \begin{pmatrix} 2 & 6 & 10 & 14 \\ 1 & 2 & 3 & 4 \end{pmatrix}, \qquad B = \begin{pmatrix} 4 & 2 & 1 & 0 \\ 14 & 10 & 6 & 2 \end{pmatrix}.$$

$$A + B = \begin{pmatrix} 2+4 & 6+2 & 10+1 & 14+0 \\ 1+14 & 2+10 & 3+6 & 4+2 \end{pmatrix} = \begin{pmatrix} 6 & 8 & 11 & 14 \\ 15 & 12 & 9 & 6 \end{pmatrix},$$

which also equals $B + A$, because the order of each addition can be reversed.

Example 2. *Associative law of addition.*

$$A = \begin{pmatrix} 3 & 4 & -2 \\ -2 & 1 & 3 \\ 1 & -3 & 2 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 1 & 1 \\ 3 & 3 & 3 \\ 5 & 5 & 5 \end{pmatrix}, \qquad C = \begin{pmatrix} 3 & 2 & -1 \\ 2 & -1 & 3 \\ 4 & 1 & -3 \end{pmatrix}.$$

$$A + B = \begin{pmatrix} 4 & 5 & -1 \\ 1 & 4 & 6 \\ 6 & 2 & 7 \end{pmatrix}, \qquad (A + B) + C = \begin{pmatrix} 7 & 7 & -2 \\ 3 & 3 & 9 \\ 10 & 3 & 4 \end{pmatrix},$$

$$B + C = \begin{pmatrix} 4 & 3 & 0 \\ 5 & 2 & 6 \\ 9 & 6 & 2 \end{pmatrix}, \qquad A + (B + C) = \begin{pmatrix} 7 & 7 & -2 \\ 3 & 3 & 9 \\ 10 & 3 & 4 \end{pmatrix}.$$

We see that $(A + B) + C = A + (B + C) = A + B + C$.

Example 3. *Multiplication by a number (*scalar*).*

$$2 \begin{pmatrix} 9 & 8 & 7 \\ 6 & 5 & 4 \\ 3 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 2\cdot 9 & 2\cdot 8 & 2\cdot 7 \\ 2\cdot 6 & 2\cdot 5 & 2\cdot 4 \\ 2\cdot 3 & 2\cdot 2 & 2\cdot 1 \end{pmatrix} = \begin{pmatrix} 18 & 16 & 14 \\ 12 & 10 & 8 \\ 6 & 4 & 2 \end{pmatrix}.$$

Example 4. *Distributive law for addition and scalar multiplication.*

$$A = \begin{pmatrix} 10 & -1 & 2 \\ 8 & -3 & 4 \\ 5 & -2 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 3 & -3 & 1 \\ -1 & 2 & 4 \\ 0 & 4 & 5 \end{pmatrix}. \qquad \text{What is } 3A + 3B\,?$$

$$3A = \begin{pmatrix} 30 & -3 & 6 \\ 24 & -9 & 12 \\ 15 & -6 & 3 \end{pmatrix}, \qquad 3B = \begin{pmatrix} 9 & -9 & 3 \\ -3 & 6 & 12 \\ 0 & 12 & 15 \end{pmatrix},$$

$$3A + 3B = \begin{pmatrix} 39 & -12 & 9 \\ 21 & -3 & 24 \\ 15 & 6 & 18 \end{pmatrix};$$

$$A + B = \begin{pmatrix} 13 & -4 & 3 \\ 7 & -1 & 8 \\ 5 & 2 & 6 \end{pmatrix}, \qquad 3(A + B) = \begin{pmatrix} 39 & -12 & 9 \\ 21 & -3 & 24 \\ 15 & 6 & 18 \end{pmatrix}.$$

So $3(A + B) = 3A + 3B$.

Example 5. *Matrix multiplication*

$$A = \begin{pmatrix} 4 & 2 & -2 \\ 0 & 3 & 5 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 2 \\ 2 & -2 \\ 2 & 3 \end{pmatrix}. \qquad \text{What are } AB \text{ and } BA\text{?}$$

$$AB = \begin{pmatrix} 4\cdot 2 + 2\cdot 2 + (-2)\cdot 2 & 4\cdot 2 + 2\cdot(-2) + (-2)\cdot 3 \\ 0\cdot 2 + 3\cdot 2 + 5\cdot 2 & 0\cdot 2 + 3\cdot(-2) + 5\cdot 3 \end{pmatrix} = \begin{pmatrix} 8 & -2 \\ 16 & 9 \end{pmatrix},$$

$$BA = \begin{pmatrix} 2\cdot 4 + 2\cdot 0 & 2\cdot 2 + 2\cdot 3 & 2\cdot(-2) + 2\cdot 5 \\ 2\cdot 4 + (-2)\cdot 0 & 2\cdot 2 + (-2)\cdot 3 & 2\cdot(-2) + (-2)\cdot 5 \\ 2\cdot 4 + 3\cdot 0 & 2\cdot 2 + 3\cdot 3 & 2\cdot(-2) + 3\cdot 5 \end{pmatrix}$$

$$= \begin{pmatrix} 8 & 10 & 6 \\ 8 & -2 & -14 \\ 8 & 13 & 11 \end{pmatrix}.$$

AB and BA both exist, but $AB \neq BA$.

Example 6. *Matrix multiplication with a different set of shapes.*

$$A = \begin{pmatrix} 4 & 5 & 6 \\ 1 & 2 & 3 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix}. \qquad \text{What are } AB \text{ and } BA\text{?}$$

$$AB = \begin{pmatrix} 4\cdot 1 + 5\cdot 1 + 6\cdot 1 & 4\cdot 1 + 5\cdot 2 + 6\cdot 3 & 4\cdot 1 + 5\cdot 4 + 6\cdot 9 \\ 1\cdot 1 + 2\cdot 1 + 3\cdot 1 & 1\cdot 1 + 2\cdot 2 + 3\cdot 3 & 1\cdot 1 + 2\cdot 4 + 3\cdot 9 \end{pmatrix} =$$

$$\begin{pmatrix} 15 & 32 & 78 \\ 6 & 14 & 36 \end{pmatrix}.$$

BA is not defined.

Example 7. *Composition of two linear transformations* (substitution of one set of linear formulas into another).

$$\begin{cases} x_1 = -y_1 + 2y_2\,, \\ x_2 = -3y_1 - y_2\,, \\ x_3 = y_1 - y_2\,; \end{cases} \qquad \begin{cases} y_1 = -z_1 - z_2 + z_3 + z_4\,, \\ y_2 = -2z_1 - z_2 - z_3 + 3z_4\,. \end{cases}$$

The matrices corresponding to these two transformations are

$$A = \begin{pmatrix} -1 & 2 \\ -3 & -1 \\ 1 & -1 \end{pmatrix}, \qquad B = \begin{pmatrix} -1 & -1 & 1 & 1 \\ -2 & -1 & -1 & 3 \end{pmatrix}.$$

The matrix of the composite function is

$$AB = \begin{pmatrix} -3 & -1 & -3 & 5 \\ 5 & 4 & -2 & -6 \\ 1 & 0 & 2 & -2 \end{pmatrix}.$$

Thus $x_1 = -3z_1 - z_2 - 3z_3 + 5z_4$, etc.

Example 8. *Associative law of multiplication.* Let

$$A = \begin{pmatrix} 1 & -1 \\ -3 & 2 \end{pmatrix}, \qquad B = \begin{pmatrix} 2 & 1 \\ 4 & 3 \end{pmatrix}, \qquad C = \begin{pmatrix} 5 & 2 \\ -1 & 0 \end{pmatrix}.$$

Then

$$AB = \begin{pmatrix} -2 & -2 \\ 2 & 3 \end{pmatrix}, \qquad (AB)C = \begin{pmatrix} -8 & -4 \\ 7 & 4 \end{pmatrix};$$

$$BC = \begin{pmatrix} 9 & 4 \\ 17 & 8 \end{pmatrix}, \qquad A(BC) = \begin{pmatrix} -8 & -4 \\ 7 & 4 \end{pmatrix}.$$

We have $(AB)C = A(BC) = ABC$.

Example 9. *Distributive laws for addition and matrix multiplication.* Let A, B, C be as in the previous example. Then

(a) $$A + B = \begin{pmatrix} 3 & 0 \\ 1 & 5 \end{pmatrix}, \qquad C(A+B) = \begin{pmatrix} 17 & 10 \\ -3 & 0 \end{pmatrix};$$

$$CA = \begin{pmatrix} -1 & -1 \\ -1 & 1 \end{pmatrix}, \qquad CB = \begin{pmatrix} 18 & 11 \\ -2 & -1 \end{pmatrix},$$

$$CA + CB = \begin{pmatrix} 17 & 10 \\ -3 & 0 \end{pmatrix}.$$

Observe that $C(A+B) = CA + CB$.

(b) $$(A+B)C = \begin{pmatrix} 15 & 6 \\ 0 & 2 \end{pmatrix}, \qquad AC = \begin{pmatrix} 6 & 2 \\ -17 & -6 \end{pmatrix},$$

$$BC = \begin{pmatrix} 9 & 4 \\ 17 & 8 \end{pmatrix}, \qquad AC + BC = \begin{pmatrix} 15 & 6 \\ 0 & 2 \end{pmatrix}.$$

Observe that $(A+B)C = AC + BC$.

Example 10. *Matrix powers.*

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{pmatrix}.$$

$$AB = \begin{pmatrix} 3 & 2 & 1 & 0 \\ 2 & 2 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad BA = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 2 & 2 \\ 0 & 1 & 2 & 3 \end{pmatrix},$$

$$A^2 = \begin{pmatrix} 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad A^3 = A^2A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$A^4 = A^3A = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix};$$

$$B^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \end{pmatrix}, \qquad B^3 = B^2B = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix},$$

$$B^4 = B^3B = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Thus $A^4 = B^4 = 0$ (the zero matrix); all higher powers are also zero.

Example 11. *A matrix polynomial.* In the notation of the previous example,

$$\begin{aligned} &1 + A + A^2 + A^3 + A^4 \\ &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\ &\quad + \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 1 & 2 & 4 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

Exercises

2.2.1 Let

$$A = \begin{pmatrix} 2 & 3 & 4 \\ 1 & 3 & 5 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & -2 & 3 \\ 5 & 1 & -3 \end{pmatrix}.$$

Calculate each of these, or declare it undefined:

(a) $A+B$, (b) $-A+2B$, (c) $3A-2B$,

(d) $(nA - B)$, where n is a given integer.

2.2.2 Let

$$A = \begin{pmatrix} 8 & 7 \\ 4 & 5 \\ 2 & 3 \\ 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \\ 7 & 8 \end{pmatrix}.$$

Calculate each of these, or declare it undefined:

(a) $-A+B$, (b) $4(-A+B)$,

(c) $-4A$, (d) $4B$, (e) $-4A+4B$,

(f) $\alpha A + \beta B$, where α and β are arbitrary real numbers.

2.2.3 Find $\begin{pmatrix} 3 & 0 & 0 & 1 \\ 2 & 3 & 1 & 0 \\ 1 & 4 & 2 & 1 \\ 0 & 1 & 3 & -2 \end{pmatrix} + \begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & 1 & 2 & 1 \\ 2 & 0 & 1 & -1 \\ 3 & 3 & 0 & 2 \end{pmatrix}.$

2.2.4 Multiply the matrices in whichever orders are possible (AB or BA):

(a) $A = \begin{pmatrix} 1 & 1 & -3 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & -1 \end{pmatrix}.$

(b) $A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 2 & -1 \\ 2 & -1 \end{pmatrix}.$

(c) $A = \begin{pmatrix} 1 & -1 & 2 \\ 2 & 3 & 4 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$

(d) $A = \begin{pmatrix} 2 & -2 \\ 4 & 5 \\ -3 & 4 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 3 \\ 2 & -1 \end{pmatrix}.$

2.2.5 Verify the associative law of matrix multiplication by calculating the products of these matrices in two ways:

(a) $$\begin{pmatrix} 7 & 2 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} -1 & 3 & 1 \\ 3 & -1 & 0 \end{pmatrix} \begin{pmatrix} 4 & 7 \\ 3 & 5 \\ 0 & 0 \end{pmatrix}.$$

(b) $$\begin{pmatrix} 2 & 3 & 1 \\ 3 & 4 & 1 \\ 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} 15 & 20 & 8 \\ -11 & -15 & -7 \\ 5 & 8 & 6 \end{pmatrix} \begin{pmatrix} -6 & 4 & 1 \\ 5 & -3 & -1 \\ -2 & 1 & 1 \end{pmatrix}.$$

2.2.6 Find AB and BA whenever they are defined.

(a) $$A = \begin{pmatrix} -2 & 3 \\ 3 & -5 \end{pmatrix}, \qquad B = \begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix}.$$

(b) $$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 9 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}.$$

(c) $$A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & -2 & 3 \end{pmatrix}, \qquad B = \begin{pmatrix} 2 & 0 \\ 1 & 3 \\ -1 & 5 \end{pmatrix}.$$

(d) $$A = \begin{pmatrix} 2 & 4 & 3 \end{pmatrix}, \quad B = \begin{pmatrix} -3 & 0 & 2 \\ 2 & -1 & 3 \\ -1 & 3 & -1 \end{pmatrix}.$$

2.2.7 Find AB when

$$A = \begin{pmatrix} 3 & 0 & 0 & 1 \\ 2 & 3 & 1 & 0 \\ 1 & 4 & 2 & 1 \\ 0 & 1 & 3 & -2 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & 1 & 2 & 1 \\ 2 & 0 & 1 & -1 \\ 3 & 3 & 0 & 2 \end{pmatrix}.$$

2.2.8 Find the *commutator* $[A, B] \equiv AB - BA$ for the matrices

(a) $$A = \begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}.$$

(b) $$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}.$$

2.2.9 Prove that for two square matrices A and B the formula $A^2 - B^2 = (A - B)(A + B)$ is true if and only if $[A, B] = AB - BA = 0$.

2.2.10 Prove that for two square matrices A and B the sum of the elements on the main diagonal of AB and BA is the same. (This number is called the *trace* of the matrix AB.).

2.2.11 Calculate these matrix powers:

(a) $\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}^3$, (b) $\begin{pmatrix} \alpha & 0 \\ \alpha & \alpha \end{pmatrix}^n$,

(c) $\begin{pmatrix} 1 & 3 & 2 \\ 1 & -1 & -1 \\ -1 & 2 & 0 \end{pmatrix}^2$, (d) $\begin{pmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & \dots & 0 \\ \multicolumn{4}{c}{\dotfill} \\ 0 & 0 & \dots & \alpha_k \end{pmatrix}^n$.

2.2.12 Find $f(A) = A^2 - A + 1$ for the matrix $A = \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}$.

2.2.13 Let $A = \begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix}$ and $f(x) = x^2 - 5x + 7$, $g(x) = x^2 - 2x + 3$. Calculate $f(A)$ and $g(A)$.

2.2.14

(a) Express $C = \begin{pmatrix} -1 & -3 \\ 3 & -4 \end{pmatrix}$ as a linear combination of

$$A = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix}.$$

(That is, find numbers r and s so that $C = rA + sB$.)

(b) Find a 2×2 matrix that cannot be expressed as a linear combination of these matrices A and B.

2.2.15

(a) To produce a boxcar load of wheat requires 3 sacks of seed and 2 tons of fertilizer. To produce a carload of milk requires 20 cows and -1 ton of fertilizer. Explain how these input requirements can be organized into a matrix, which can be used to calculate the inputs needed to produce x loads of wheat and y loads of milk.

(b) To produce 1 million biscuits requires 2 loads of wheat and 1 load of milk. Show how to use matrices to calculate the seed, cows, and fertilizer needed to produce 20 million biscuits.

2.2.16 The Aggie Industrial Engineering Demonstration Factory manufactures zarfs and bibcocks. A zarf requires 1 kilogram each of steel and aluminum. A bibcock takes 2 kilos of steel and 3 of aluminum. Steel costs \$2 per kilo, while aluminum is \$5 per kilo. Show how to organize these facts into matrices and thereby obtain the matrix of the linear function telling us the cost of z zarfs and b bibcocks.

2.2.17 Producing a car requires 1 ton of steel, 1 ton of aluminum, and 10 pounds of glass. Producing an airplane requires 2 tons of steel, 3 tons of aluminum, and 60 pounds of glass. Steel costs \$500 per ton, aluminum \$1,000 per ton, and glass \$10 per pound. Organize these facts into matrices, and find the matrix that tells you how much money is needed for the raw materials to make c cars and a airplanes.

2.2.18 To produce a zingabob requires 5 pounds of krypton plastic and 30 cubic feet of steam. To produce a pound of krypton plastic requires 10 pounds of kryptonite and 50 hours of labor. To produce a cubic foot of steam requires an ounce of water and an hour of labor. Show how these input requirements can be organized into matrices, and how the matrices can be used to calculate the amount of kryptonite, water, and labor needed to produce a certain number of zingabobs.

2.2.19 Col. Roger Rapidrudder now has a desk job at the Pentagon supervising weapons production.* A howitzer contains 1000 cubic inches of steel and carries an allowance of 100 cubic inches of gunpowder. A rifle contains 5 cubic inches of steel and is allocated 2 cubic inches of gunpowder. Steel weighs 2 pounds per cubic inch and costs \$3 per cubic inch. Gunpowder weighs 0.1 pound per cubic inch and costs \$4 per cubic inch. Remind Roger of how to organize these facts into matrices, and find the matrix which should be used to calculate the total weight w and total cost c of the material needed to make and support h howitzers and r rifles.

2.2.20 Producing a car requires 1 ton of steel and 0.5 ton of plastic. Producing an airplane requires 5 tons of steel and 2 tons of plastic. Producing a ton of steel consumes 3 tons of bituminous coal and 20 barrels of water. Producing a ton of plastic consumes 2 tons of coal and 50 barrels of water. Organize these facts into matrices, and find the matrix that tells you how much coal (b) and water (w) is needed to make c cars and a airplanes.

2.2.21 Let $A = \begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix}, \quad B = \begin{pmatrix} -2 & 2 & 2 \\ 1 & 0 & 1 \end{pmatrix}.$

(a) Calculate all of the following that are defined: $AB,\ BA,\ A^{\mathrm{t}}$.

---

* Roger's exciting adventures in his younger days will figure prominently in our later examples and exercises.

(b) Use these matrices to make up an economics word problem of the type "To produce a zorch requires so many sacks of freebles and … ". (Invent two sets of industries, the outputs of one being the raw materials of the other.)

2.2.22 Prove

(a) $(A^{\mathrm{t}})^{\mathrm{t}} = A$, (b) $(A + B)^{\mathrm{t}} = A^{\mathrm{t}} + B^{\mathrm{t}}$,

(c) $(AB)^{\mathrm{t}} = B^{\mathrm{t}}A^{\mathrm{t}}$.

2.2.23 Decompose $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 2 \\ -6 & 0 & 9 \end{pmatrix}$ into its symmetric and antisymmetric parts.

2.2.24 Prove that the decomposition of a square matrix into its symmetric and antisymmetric parts is unique. HINT: First show that the only matrix that is both symmetric and antisymmetric is the zero matrix.

2.2.25 Prove that if A is antisymmetric, then its diagonal elements, A_{11}, A_{22} …, are all 0.

2.2.26 If $C = \begin{pmatrix} 3 & 2-6i \\ 2+6i & 1 \end{pmatrix}$ (where $i^2 = -1$), calculate C^2.

2.2.27 To make a loaf of bread, the bakery uses 2 cups of flour and 1 cup of sugar. To make a pie, it uses 1 cup of flour and 3 cups of sugar. Sugar costs 3 cents per cup, flour costs 2 cents per cup. In addition, there is a sales tax of 1 cent per cup on both commodities. Show how to organize the facts into matrices, and find the matrix that should be used to calculate how much the bakery must pay the suppliers and how much it must pay the government if it buys ingredients for b loaves of bread and p pies.

2.3 Inverses

Definition: If A is a square (i.e., $n \times n$) matrix, then B (which must also be $n \times n$) is the *inverse* of A if

$$AB = I = BA.$$

The motivation for this definition again comes the interpretation of matrices in terms of linear systems. Consider the system of equations $\vec{y} =$

$A\vec{x}$, where $\vec{y}$ and $\vec{x}$ are column vectors $(n \times 1)$ and A is square $(n \times n)$. If A has an inverse, B, then $\vec{x} = B\vec{y}$ is a solution of the system (in fact, the only solution).

PROOF: $A\vec{x} = AB\vec{y} = I\vec{y} = \vec{y}$. Conversely, if $\vec{y} = A\vec{x}$, then $B\vec{y} = BA\vec{x} = I\vec{x} = \vec{x}$.

FACTS AND REMARKS ABOUT INVERSES

(1) If A is *not* square, then $AB = I$ and $BA = I$ *can't* both be true for any B. So nonsquare matrices don't have inverses. Here's an example which shows this:

$$A = \begin{pmatrix} a \\ b \end{pmatrix}, \qquad B = \begin{pmatrix} c & d \end{pmatrix}.$$

$$AB = \begin{pmatrix} a \\ b \end{pmatrix} \begin{pmatrix} c & d \end{pmatrix} = \begin{pmatrix} ac & ad \\ bc & bd \end{pmatrix} \overset{?}{=} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

If $ad = bc = 0$, then

$$\text{either} \quad a = 0 \quad \text{or} \quad d = 0,$$

and

$$\text{either} \quad b = 0 \quad \text{or} \quad c = 0.$$

Thus the required condition $ac = bd = 1$ is impossible. (Later, when we discuss the *rank* of matrices, we'll understand on a more fundamental level why this had to happen.)

(2) If A is square and has an inverse B, then that inverse is unique. PROOF:

$$B_1 A = I = AB_2 \;\Rightarrow\; B_1 = B_1 I = B_1 A B_2 = I B_2 = B_2\,.$$

Now that we know it is well-defined, we can give the inverse B the notation A^{-1}. (This notation is reasonable since A^{-1} acts like the reciprocal of A with respect to matrix multiplication. Furthermore, a 1×1 matrix is just a number, and in that case the inverse is exactly the same thing as the reciprocal.)

(3) If A is *square* and there is a B satisfying $BA = I$, then $AB = I$ also; so $B = A^{-1}$ and is unique. Similarly, $AB = I$ implies $BA = I$.

(See Exercise 5.4.13 for a proof.) This observation is very useful in checking a calculation of A^{-1}: we need only work out one matrix multiplication, not both, to be confident that we have found the right answer.

On the other hand, if A is *not square*, there may be many *left inverses* (B's satisfying $BA = I$) and no *right inverses* (satisfying $AB = I$); or many right inverses and no left inverses; or none of either.

(4) If A has no inverse, it is called *singular*. The reduced (row echelon) form of a nonsingular (or invertible) matrix is the identity matrix. The reduced form of a singular square matrix has at least one row of zeros at the bottom, as in these example patterns:

$$\begin{pmatrix} 1 & 5 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 & \pi \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Note in each case how the row reduction process has ground to a halt.

Algebraic Identities 4:

$$(A^{-1})^{-1} = A.$$

$$(ABC\cdots)^{-1} = \cdots C^{-1}B^{-1}A^{-1} \quad \text{(if the inverses on the right exist).}$$

$$(A^{\mathrm{t}})^{-1} = (A^{-1})^{\mathrm{t}} \quad \text{(if either inverse exists).}$$

$$(rA)^{-1} = \frac{1}{r}\,A^{-1} \quad \text{(if } r \neq 0 \text{ and } A^{-1} \text{ exists).}$$

On the other hand, of course $(A + B)^{-1} \neq A^{-1} + B^{-1}$. (You wouldn't do that for numbers, would you?)

Finding the inverse by row reduction

Algorithm: To find the inverse of a square matrix A, form the huge augmented matrix

$$(\,A \mid I\,) = \left(\begin{array}{cccc|cccc} a_{11} & a_{12} & \dots & a_{1n} & 1 & 0 & \dots & 0 \\ a_{21} & & & & 0 & 1 & & \\ \vdots & & \ddots & & \vdots & & \ddots & \\ a_{n1} & & & a_{nn} & 0 & & & 1 \end{array}\right).$$

Reduce it. If A is nonsingular, you will get

$$\left(\begin{array}{ccc|ccc} 1 & 0 & \dots & b_{11} & b_{12} & \dots \\ 0 & 1 & \dots & b_{21} & & \\ \vdots & & \ddots & \dots & & \end{array}\right) = (\, I \mid B \,),$$

and this B will equal A^{-1}! On the other hand, if A is singular, then you will eventually run into a situation like

$$\left(\begin{array}{ccc|cc} 1 & 7 & 8 & \dots & \dots \\ 0 & 0 & 9 & \dots & \dots \\ 0 & 0 & 5 & \dots & \dots \end{array}\right).$$

Then stop! The reduced form of the left half is not going to be the identity. There is no point in finishing the reduction, since you now know that A has no inverse.

Example 1: $A = \begin{pmatrix} 1 & 2 \\ 1 & 4 \end{pmatrix}.$

$$\left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 1 & 4 & 0 & 1 \end{array}\right) \longrightarrow \begin{pmatrix} 1 & 2 & 1 & 0 \\ 0 & 2 & -1 & 1 \end{pmatrix}$$

$$\longrightarrow \left(\begin{array}{cc|cc} 1 & 0 & 2 & -1 \\ 0 & 1 & -\frac{1}{2} & \frac{1}{2} \end{array}\right).$$

Thus

$$A^{-1} = \begin{pmatrix} 2 & -1 \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Check: $AA^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$

Example 2: $A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}.$

$$\left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 2 & 4 & 0 & 1 \end{array}\right) \longrightarrow \begin{pmatrix} 1 & 2 & 1 & 0 \\ 0 & 0 & -2 & 1 \end{pmatrix}$$

and we see that A is singular.

Why the algorithm works: If $B = A^{-1}$, then B satisfies the equation $AB = I$ — that is,

$$\begin{pmatrix} A_{11} & A_{12} & \dots \\ A_{21} & \dots & \\ \dots & & \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} & \dots \\ B_{21} & \dots & \\ \dots & & \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \vdots \\ 0 & \dots & \ddots \end{pmatrix}.$$

By the definition of matrix multiplication, therefore, the jth column of B is the solution of the system

$$A\begin{pmatrix} B_{1j} \\ B_{2j} \\ \vdots \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \equiv \vec{e}_j\,,$$

where $\vec{e}_j$ is the vector with a 1 in the jth row and zeros everywhere else. We could solve this system by reducing the augmented matrix $(\,A \mid \vec{e}_j\,)$, getting

$$\left(\begin{array}{c|c} I & \begin{smallmatrix} B_{1j} \\ \vdots \end{smallmatrix} \end{array}\right).$$

By reducing $(\,A \mid I\,)$, we are doing this *for all the columns* $\vec{e}_j$ *at once*, automatically getting the answers stacked together in the right order to constitute A^{-1}.

More examples

Example 3.

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 2 \\ 1 & 2 & 1 \end{pmatrix}.$$

We find the inverse by row reduction. $(\,A \mid I\,)$ is

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 1 & 2 & 2 & 0 & 1 & 0 \\ 1 & 2 & 1 & 0 & 0 & 1 \end{array}\right) \xrightarrow[]{\substack{(2)-(1)\to(2) \\ (3)-(1)\to(3)}} \left(\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{(3)-(2)\to(3)} \left(\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 0 & -1 & 1 \end{array}\right) \xrightarrow{(1)+2(3)\to(1)}$$

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & -2 & 2 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 0 & -1 & 1 \end{array}\right) \xrightarrow{(3)\to-(3)} \left(\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & -2 & 2 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & -1 \end{array}\right)$$

$$\xrightarrow{(1)-(2)\to(1)} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 2 & -3 & 2 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & -1 \end{array}\right) = \left(\,I \mid A^{-1}\,\right).$$

Thus

$$A^{-1} = \begin{pmatrix} 2 & -3 & 2 \\ -1 & 1 & 0 \\ 0 & 1 & -1 \end{pmatrix}.$$

Example 4.

$$A = \begin{pmatrix} 2 & 7 & 15 \\ 1 & 3 & 6 \\ 3 & 10 & 21 \end{pmatrix}.$$

$$(\,A \mid I\,) = \left(\begin{array}{ccc|ccc} 2 & 7 & 15 & 1 & 0 & 0 \\ 1 & 3 & 6 & 0 & 1 & 0 \\ 3 & 10 & 21 & 0 & 0 & 1 \end{array}\right) \xrightarrow{(2) \leftrightarrow (1)} \left(\begin{array}{ccc|ccc} 1 & 3 & 6 & 0 & 1 & 0 \\ 2 & 7 & 15 & 1 & 0 & 0 \\ 3 & 10 & 21 & 0 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{\substack{(2)-2(1)\to(2) \\ (3)-3(1)\to(3)}} \left(\begin{array}{ccc|ccc} 1 & 3 & 6 & 0 & 1 & 0 \\ 0 & 1 & 3 & 1 & -2 & 0 \\ 0 & 1 & 3 & 0 & -3 & 1 \end{array}\right) \xrightarrow{(3)-(2)\to(3)}$$

$$\left(\begin{array}{ccc|ccc} 1 & 3 & 6 & 0 & 1 & 0 \\ 0 & 1 & 3 & 1 & -2 & 0 \\ 0 & 0 & 0 & -1 & -1 & 1 \end{array}\right).$$

Thus A is singular.

Example 5.

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

$$\left(\begin{array}{cccc|cccc} 1 & 2 & 3 & 4 & 1 & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right) \xrightarrow{\substack{(3)-2(4)\to(3) \\ (2)-3(4)\to(2) \\ (1)-4(4)\to(1)}}$$

$$\left(\begin{array}{cccc|cccc} 1 & 2 & 3 & 0 & 1 & 0 & 0 & -4 \\ 0 & 1 & 2 & 0 & 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right) \xrightarrow{\substack{(1)-3(3)\to(1) \\ (2)-2(3)\to(2)}}$$

$$\left(\begin{array}{cccc|cccc} 1 & 2 & 0 & 0 & 1 & 0 & -3 & 2 \\ 0 & 1 & 0 & 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right) \xrightarrow{(1)-2(2)\to(1)}$$

$$\left(\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right)$$

Thus

$$A^{-1} = \begin{pmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Example 6. Consider the most general 2×2 matrix, $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. This problem is actually best solved by determinants (see Sec. 2.5), but here we shall see what row reduction has to offer. We will be led naturally to the *determinant*

$$\Delta = \det(A) = ad - bc$$

and to the essential condition $\Delta \neq 0$. To avoid considering a row interchange, we must assume that $a \neq 0$.

$$(\,A \mid I\,) = \left(\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right) \xrightarrow{(2) - \frac{c}{a}(1) \to (2)} \left(\begin{array}{cc|cc} a & b & 1 & 0 \\ 0 & d - \frac{bc}{a} & -\frac{c}{a} & 1 \end{array}\right)$$

$$\xrightarrow{(1) \to \frac{1}{a}(1)} \left(\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ 0 & \frac{ad-bc}{a} & -\frac{c}{a} & 1 \end{array}\right) \xrightarrow{(2) \to \frac{a}{\Delta}(2)} \left(\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ 0 & 1 & -\frac{c}{\Delta} & \frac{a}{\Delta} \end{array}\right)$$

$$\xrightarrow{(1) - \frac{b}{a}(2) \to (1)} \left(\begin{array}{cc|cc} 1 & 0 & \frac{1}{a} + \frac{bc}{a\Delta} & -\frac{b}{\Delta} \\ 0 & 1 & -\frac{c}{\Delta} & \frac{a}{\Delta} \end{array}\right) = (\,I \mid A^{-1}\,).$$

Thus

$$A^{-1} = \begin{pmatrix} \frac{d}{\Delta} & -\frac{b}{\Delta} \\ -\frac{c}{\Delta} & \frac{a}{\Delta} \end{pmatrix} = \frac{1}{\Delta}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Check:

$$AA^{-1} = \frac{1}{\Delta}\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \frac{1}{\Delta}\begin{pmatrix} \Delta & 0 \\ 0 & \Delta \end{pmatrix} = I.$$

We leave the case $a = 0$ to the reader (Exercise 2.3.1).

Example of the example: $A = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$. We calculate

$$\Delta = \det(A) = \cos^2\alpha + \sin^2\alpha = 1 \neq 0.$$

In the notation of Example 6,

$$a = \cos\alpha, \quad b = -\sin\alpha, \quad c = \sin\alpha, \quad d = \cos\alpha.$$

So

$$A^{-1} = \frac{1}{\Delta}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}.$$

REMARK: A represents a rotation of the plane through the angle α; A^{-1} represents the rotation through $-\alpha$ — i.e., through the same angle in the opposite direction.

Example 7.

$$A = \begin{pmatrix} a_1 & 0 & 0 & \dots & 0 \\ 0 & a_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_n \end{pmatrix}, \qquad a_i \neq 0, \quad i = 1, 2, \dots, n.$$

$$(\, A \mid I \,) = \left(\begin{array}{ccccc|ccccc} a_1 & 0 & 0 & \dots & 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & a_2 & 0 & \dots & 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_n & 0 & 0 & 0 & \dots & 1 \end{array}\right).$$

Multiply successively the first row by $\frac{1}{a_1}$, the second by $\frac{1}{a_2}$, $\dots$, the nth (last) row by $\frac{1}{a_n}$. The result is

$$\left(\begin{array}{ccccc|ccccc} 1 & 0 & 0 & \dots & 0 & \frac{1}{a_1} & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 & \frac{1}{a_2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 & 0 & 0 & \dots & \frac{1}{a_n} \end{array}\right) = (\, I \mid A^{-1} \,).$$

Thus

$$A^{-1} = \begin{pmatrix} \frac{1}{a_1} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{a_2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{a_n} \end{pmatrix}.$$

In other words, *the inverse of a diagonal matrix is the diagonal matrix formed from the reciprocals of the diagonal elements of the original matrix.* Once this has been pointed out, it is quite obvious, and it should not be necessary to go through the row reduction again in such a case.

Remark: In fact, it is easy to see (by mentally carrying out the checking multiplication, $AA^{-1} \stackrel{?}{=} I$) that a similar fact is true for any *block-diagonal matrix*; that is, a square matrix whose only nonzero elements are concentrated into square blocks along the main diagonal. In such a case, the inverse of the large matrix is obtained simply by replacing each of the blocks with its inverse. For instance, in view of Example 6, we have

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin \\ 0 & -\sin\alpha & \cos\alpha \end{pmatrix}.$$

These matrices represent rotations about the x axis in three-dimensional space with the usual coordinates, (x, y, z).

Exercises

2.3.1 Show that the conclusion of Exercise 6 remains valid when $a = 0$, provided that $\Delta \neq 0$.

2.3.2 Find the inverses of these matrices, if they exist.

(a) $\begin{pmatrix} 0 & 4 \\ 1 & 2 \end{pmatrix}$, (b) $\begin{pmatrix} 1 & 2 \\ 0 & 4 \end{pmatrix}$, (c) $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$.

2.3.3 Find the inverses of these matrices (if they exist).

(a) $\begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & -1 \\ 2 & 2 & 2 \end{pmatrix}$, (b) $\begin{pmatrix} 2 & 1 & 7 \\ 1 & 1 & 1 \\ -2 & 0 & -10 \end{pmatrix}$.

2.3.4 Does the matrix $M = \begin{pmatrix} 1 & 3 & -1 \\ 1 & -1 & 1 \\ 3 & 1 & 1 \end{pmatrix}$ have an inverse? Explain. [Hint: Compare Exercise 2.1.4(c).]

2.3.5 Find the inverses of these matrices, if they exist.

(a) $\begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix}$, (b) $\begin{pmatrix} 2 & 3 \\ 5 & 8 \end{pmatrix}$, (c) $\begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix}$.

2.3.6 Find

(a) $\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ 2 & 2 & -1 \end{pmatrix}^{-1}$, (b) $\begin{pmatrix} 1 & 0 & 2 \\ -1 & 1 & 2 \\ 3 & -2 & -2 \end{pmatrix}^{-1}$.

In the next seven exercises, calculate the indicated inverses (if they exist).

2.3.7 $\begin{pmatrix} -1 & 2 & 3 \\ 2 & 1 & 5 \\ 4 & -6 & -7 \end{pmatrix}^{-1}$

2.3.8 $\begin{pmatrix} 2 & 6 & 5 \\ 5 & 3 & -2 \\ 7 & 4 & -3 \end{pmatrix}^{-1}$

2.3.9 $\begin{pmatrix} 2 & 5 & -1 \\ -1 & 3 & -2 \\ 0 & -6 & 3 \end{pmatrix}^{-1}$

2.3.10 $\begin{pmatrix} 2 & 0 & 3 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}^{-1}$

2.3.11 $\begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & -2 \\ 2 & -2 & -6 \end{pmatrix}^{-1}$

2.3.12 $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 2 & 1 & 0 \\ 4 & 3 & 2 & 1 \end{pmatrix}^{-1}$

2.3.13 $\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}^{-1}$

2.3.14 Solve these matrix equations for the unknown matrix, X or Y.

(a) $\begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix} X = \begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix}$

(b) $Y \begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix}$

(c) $X \begin{pmatrix} 3 & 1 \\ 5 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$

2.3.15 Solve these matrix equations for the unknown matrix, X or Y.

(a) $X \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ 2 & 2 & -1 \end{pmatrix} = \begin{pmatrix} 3 & 2 & 4 \\ -1 & 2 & 1 \\ 5 & 3 & 2 \end{pmatrix}$

(b) $\begin{pmatrix} 3 & 1 \\ 5 & 2 \end{pmatrix} Y \begin{pmatrix} 4 & 5 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$

2.3.16 Prove that $(AB)^{-1} = B^{-1}A^{-1}$ (assuming that the inverses on the right exist).

2.3.17 Prove (assuming no existence problems)

(a) $(A^{-1})^{-1} = A$, (b) $(A^{\mathrm{t}})^{-1} = (A^{-1})^{\mathrm{t}}$.

2.3.18 Calculate and simplify (using $e^{i\theta} = \cos\theta + i\sin\theta$)

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} e^{2i} & 0 \\ 0 & e^{-2i} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

2.4 Functions and Gradient Vectors

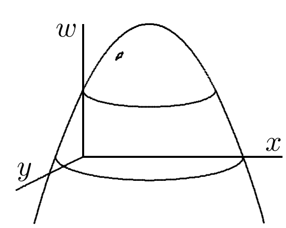
In Sec. 1.4 we looked at vector-valued functions of a real variable (alias "curves") and their tangent vectors. Now we shall turn to the reverse situation, a real-valued function of a vector variable:

$$f\colon \mathbf{R}^n \to \mathbf{R}.$$

As in the other case, there are (at least) two different ways in which one can visualize such a function geometrically.

1. If $n = 2$, the *graph* of the function is a *surface*:

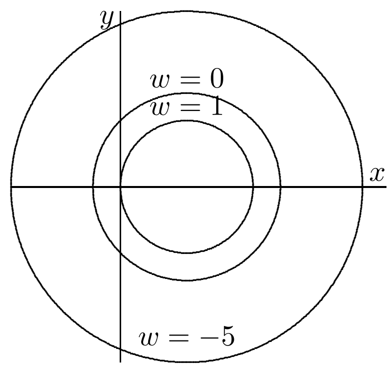
$$w = 2 - (x-1)^2 - y^2.$$



In the general case, the graph is an n-dimensional *hypersurface* in an $(n+1)$-dimensional space.

2. For our example of a function of two independent variables, in $\mathbf{R}^2$ we can draw *level curves* (also known as *contour lines*) along which the

function is constant.



When $n = 3$ these become *level surfaces*, and so on; a convenient general term is *level sets*. (Note that the level curves could "fatten out", if the function were constant over a two-dimensional region. In the most typical situation, however, the level hypersurfaces will have dimension $n - 1$, because the equation

$$f(x_1, \ldots, x_n) = w = \text{constant}$$

places one constraint on the n variables.)

The functions we are considering now are, generally speaking, not linear. However, if a function is sufficiently "smooth", then its "local" behavior near a point $\vec{x}_0$ in $\mathbf{R}^n$ can be described in the languages of linear functions and matrices. This is the central idea of differential calculus, boosted to n-dimensional space.

From third-semester calculus you know how to calculate *partial derivatives* with respect to each variable (in the example, $\frac{\partial w}{\partial x}$ and $\frac{\partial w}{\partial y}$). Evaluating these n functions at $\vec{x}_0$ we get n numbers, which we can put together into a row vector,

$$\nabla f(\vec{x}_0) \equiv \left(\left.\frac{\partial f}{\partial x_1}\right|_{\vec{x}_0}, \left.\frac{\partial f}{\partial x_2}\right|_{\vec{x}_0}, \ldots \right).$$

This vector is called the *gradient* of f at that point.

Now consider the *linear* function

$$w = f(\vec{x}_0) + (\vec{x} - \vec{x}_0) \cdot \nabla f(\vec{x}_0). \tag{2}$$

Note the close analogy between this equation and equation (1) in Sec. 1.4. When $n = 2$, (2) can be written out as

$$w = f(x_0, y_0) + \frac{\partial f}{\partial x}\,(x - x_0) + \frac{\partial f}{\partial y}\,(y - y_0)$$

(where the partial derivatives are understood to be evaluated at $x = x_0$, $y = y_0$), which is recognized as the equation of the *tangent plane* to the graph of f at the point $\vec{x}_0$. Of all the planes (*flat* surfaces) through $\vec{x}_0$, this one lies closest to the graph (a curved, but smooth, surface). In more numerical terms, for $\vec{x}$ near $\vec{x}_0$ the formula (2) is the best way to approximate $f(x)$ by a linear function. For general n, (2) gives the best approximation to the graph near $\vec{x}_0$ by a *hyperplane* (a flat n-dimensional hypersurface). In the next chapter we give a definition of tangent hyperplanes that does not depend on geometrical intuition (which becomes harder and harder to rely on as the number of variables increases).

Remember that for $\vec{v} \in \mathbf{R}^n$, the *directional derivative of f at $\vec{x}_0$ along $\vec{v}$* is defined as the rate of change of f along the line through $\vec{x}_0$ parallel to $\vec{v}$:

$$\frac{\partial f}{\partial \vec{v}}(\vec{x}_0) \equiv \lim_{h\to 0} \frac{f(\vec{x}_0 + h\vec{v}) - f(\vec{x}_0)}{h}.$$

This number depends linearly on $\vec{v}$ (more on this in a moment). However, it is customary to restrict attention to *unit vectors*. To every $\vec{v}$ corresponds a unit vector,

$$\hat{v} \equiv \frac{\vec{v}}{\|\vec{v}\|}.$$

If $\|\vec{v}\| \neq 1$, "the directional derivative in the direction of $\vec{v}$" means $\frac{\partial f}{\partial \hat{v}}$, not $\frac{\partial f}{\partial \vec{v}}$. (This quantity is also called the "rate of change of f in the direction of $\vec{v}$".)

What is the relation of the directional derivative to partial derivatives and the gradient?

Theorem 1:

(1) If $\vec{v} = \vec{e}_j$, the unit vector along the x_j axis, then the directional derivative is the partial derivative:

$$\frac{\partial f}{\partial \vec{e}_j} = \frac{\partial f}{\partial x_j}.$$

(2) In general,

$$\frac{\partial f}{\partial \vec{v}}(\vec{x}) = \sum_{j=1}^{n} \left.\frac{\partial f}{\partial x_j}\right|_{\vec{x}} v_j.$$

Thus any directional derivative is a linear combination of partial derivatives. Moreover, it can be written as the dot product of the direction vector with the gradient vector:

$$\frac{\partial f}{\partial \hat{v}} = \hat{v} \cdot \nabla f. \tag{$*$}$$

The following important facts, which give geometrical significance to the gradient, are consequences of $(*)$.

Theorem 2:

(1) $\nabla f(\vec{x})$ points in the *direction of fastest increase* of f (starting out from $\vec{x}$). This maximum rate of increase equals $\|\nabla f\|$. (The direction of $-\nabla f$ is the direction of fastest *decrease*.)

(2) If $\nabla f(\vec{x}) \neq 0$, then $\nabla f(\vec{x})$ is *perpendicular to the level set* of f through $\vec{x}$ (that is, the set of points $\vec{y}$ such that $f(\vec{y}) = f(\vec{x})$). As previously remarked, the level set is then a curve in dimension 2, a surface in dimension 3, or an $n-1$-dimensional hypersurface in dimension n. On the other hand, if $\nabla f(\vec{x}) = 0$, the level set through $\vec{x}$ may not be a hypersurface, and even if it is, the statement that ∇f is perpendicular to it has no content.

Proof: (1) Use the well known relation for the dot product,

$$\vec{u} \cdot \vec{v} = \|\vec{u}\| \|\vec{v}\| \cos\theta$$

where θ is the angle between $\vec{u}$ and $\vec{v}$. From this it is clear that $(*)$ attains its maximum value when $\vec{v}$ is parallel to ∇f, and that this value is $\|\nabla f\|$ when $\vec{v}$ is a unit vector (thus defining a rate of increase or decrease). (2) The rate of change of f as we move through $\vec{x}$ along a curve lying in the level surface is $(*)$, with $\vec{v}$ being the tangent vector to the curve. But that rate of change is zero, by definition of a level surface. Therefore, ∇f is perpendicular to all such tangent vectors, which is what it means to be perpendicular to the surface. (We have used 3-dimensional terminology for vividness, but the same argument applies in all dimensions.)

The chain rule

Another of the many things you're expected to remember from your third-semester calculus course is the extension of the chain rule to functions of several variables. For a review and first application of this, let's consider this vignette:

> While flying northeast at $1000\sqrt{2}$ feet per second, Roger Rapidrudder measured the gradient vector of the air temperature to be
>
> $$(0.0095, 0.0023, -0.0196) \qquad \text{in units of degrees per foot.}$$
>
> (The coordinate axes point east, north, and up, in that order.) How fast (in degrees per second) was the temperature outside the plane changing?

SOLUTION: Let $\vec{x}(t)$ be the plane's trajectory, and let $T(\vec{z})$ be the temperature at $\vec{z}$. Then

$$\frac{dT}{dt} = \sum_{j=1}^{3} \left.\frac{\partial T}{\partial z_j}\right|_{\vec{z}=\vec{x}(t)} \frac{dx_j}{dt} = \nabla T(\vec{x}) \cdot \vec{x}\,'(t).$$

Roger has velocity vector

$$\vec{v}(t) = \vec{x}\,'(t) = \begin{pmatrix} 1000 \\ 1000 \\ 0 \end{pmatrix}$$

at all t, hence $dT/dt = 9.5 + 2.3 = 11.8$. (This rate of change as measured by Roger in motion should not be confused with the abstract mathematical rate of change discussed earlier; that one is $\hat{v} \cdot \nabla T$, hence smaller than dT/dt by a factor $1000\sqrt{2}$, and it has units of degrees per foot, not per second.)

Note that people who fully understand what they're doing are licensed to drop the distinction between $\vec{z}$ and $\vec{x}$ and write the formula as

$$\frac{dT}{dt} = \sum_{j=1}^{3} \frac{\partial T}{\partial x_j} \frac{dx_j}{dt}\,.$$

In short, the formula looks like the single-variable chain rule, except that we write one term for each component of the intermediate vector variable and sum them up. The vectorial form $T' = \nabla T \cdot \vec{x}\,'$ is written as a dot product, but it can also be seen as the matrix product of the row matrix ∇T and the column matrix $\vec{x}\,'$. Indeed, the latter interpretation is the more fundamental one, and the one that will be extended in the next chapter to situations where the initial and final variables also are multidimensional.

As a down payment, we observe that a (differentiable) function $\vec{f}$ from $\mathbf{R}^n$ into $\mathbf{R}^p$ is associated with a very important matrix-valued function, the matrix of its partial derivatives:

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots \\ \frac{\partial f_2}{\partial x_1} & \cdots & \\ \vdots & & \end{pmatrix}.$$

In Sec. 3.4 we will interpret this as the matrix of a certain linear transformation, thus attaining a more profound understanding of what partial derivatives are.

Let us consider some applications of the chain rule:

Example 1. The Leibnitz rule (product rule) is a special case of the multivariable chain rule (although, of course, it can't be presented to first-semester calculus students in that way). Consider the problem of evaluating

$$\frac{d}{dx}[f(x)g(x)].$$

Define $\vec{F}\colon \mathbf{R} \to \mathbf{R}^2$ and $G\colon \mathbf{R}^2 \to \mathbf{R}$ by

$$\vec{F}(x) \equiv \begin{pmatrix} x \\ x \end{pmatrix}, \qquad G(\vec{y}) \equiv f(y_1)g(y_2).$$

Note that since F is a linear function, its best linear approximation is itself; its tangent vector is the same as its coefficient matrix. Thus

$$\begin{aligned} \frac{d}{dx}[f(x)g(x)] &= \frac{d}{dx}G\bigl(F(x)\bigr) \\ &= \nabla G\bigl(F(x)\bigr) \cdot \vec{F}'(x) \\ &= \begin{pmatrix} f'(y_1)g(y_2) & f(y_1)g'(y_2) \end{pmatrix}\Big|_{\vec{y}=F(x)} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= f'(x)g(x) + f(x)g'(x). \end{aligned}$$

Example 2. The same argument works for *any* function which depends on x in "more than one place". Every place we see an x, we differentiate with respect to it; then we add up all the resulting terms. For example, by the "first fundamental theorem of calculus",

$$\begin{aligned} \frac{d}{dx}\int_0^x (x+t)\,dt &= (x+t)\big|_{t=x} + \int_0^x \frac{\partial}{\partial x}(x+t)\,dt \\ &= 2x + \int_0^x dt \\ &= 3x. \end{aligned}$$

In this example we have

$$G(\vec{y}) \equiv \int_0^{y_1} (y_2 + t)\, dt$$

(and the same $\vec{F}$ as in Example 1).

Exercises

2.4.1 Let $f(x, y) = \frac{1}{4}(x + y^2)$.

(a) Calculate the gradient, ∇f.

(b) Sketch the gradient vector at some representative points. Also, sketch (and label) some level curves of the function f.

2.4.2 Let $f(x, y) = \frac{1}{4}x^2 + y^2$. Sketch the level curve of f that passes through the point $(x, y) = (0, 1)$, and sketch the gradient vector of f at 4 representative points on that curve.

2.4.3 The thickness of a square aluminum plate (in hundredths of a millimeter) is given by the formula

$$\rho(x, y) = 2x + 3y^2 \quad (1 < x < 2, \quad 1 < y < 2).$$

(a) Sketch the curve on which the thickness equals 6 units.

(b) Find the direction of steepest increase of ρ at the center of the plate.

(c) Find the rate of change of ρ at the center of the plate as the sample point moves toward the corner point (1,1).

2.4.4 Let $u(x, y) = \exp\left(\dfrac{x}{x^2 + y^2}\right)$.

(a) Sketch the level curves $u = e^2$ and $u = e^{1/4}$.

(b) Calculate and draw the vector ∇u at the points $(1, -1)$, $(1, 0)$, and $(-1, 1)$.

2.4.5 Find the gradient vector field of the function $f(\vec{r}) = -xyz + y^2z + z^2$, and use this information to construct the tangent plane to the surface $f = 1$ at the point $(1, 1, 1)$. (Use Theorem 2(2).)

2.4.6 A temperature field is given by the function $T(\vec{r}) = x^2 - 2xy + 2y^2$.

(a) In what direction does the temperature have maximal growth at the point $\vec{r} = (-1, 1, 0)$?

(b) What is the rate of change of temperature (with respect to time) along the path

$$\vec{r}(t) = \begin{pmatrix} t \\ t^2 \\ 1 - t^2 \end{pmatrix}$$

at time $t = -1$?

2.4.7 Is the plane $z = 0$ a tangent plane at the point $(x, y, z) = (0, 0, 0)$ to:

(a) the elliptic paraboloid $z = 2x^2 + 3y^2$?

(b) the cone $z = \sqrt{x^2 + 2y^2}$?

(c) the hyperbolic paraboloid $z = 9xy$?

2.4.8 In what direction must the point (x, y, z) move when passing through the location $(1, -1, 1)$ in order that the function

$$f(x, y, z) = \frac{x}{y} + \frac{y}{z} + \frac{z}{x}$$

grow with the maximal speed?

2.4.9 Find the gradient of the function $u = x^3 + 2y^3 + 3z^3 - 4xyz$, and determine at which points it

(a) is parallel to the x axis;

(b) is perpendicular to the direction of the line $y = x$;

(c) is equal to 0.

2.4.10 Find the gradient of

$$u = \frac{2x}{x^2 - y^2 + z^2}\,.$$

2.4.11 Find the tangent plane to the graph of $z = x^3 - y^3$ at $(x, y) = (1, -1)$.

2.4.12 Find the tangent plane to $z = \cos\frac{x}{y}$ at $(x, y) = \left(1, \frac{3}{\pi}\right)$.

2.4.13 The graph of a function $z = f(x, y)$ can be regarded as a level curve of the three-variable function $g(x, y, z)$. Find a formula for the tangent plane at an arbitrary point on the graph by means of Theorem 2(2), and verify that your formula agrees with (2).

2.4.14 Calculate $\dfrac{d}{dx}\displaystyle\int_{-\infty}^{x}(1+xt)\,e^{-t^2}\,dt$, and then evaluate the limit of your answer as $x \to +\infty$.

2.4.15 Use the chain rule to calculate $\dfrac{d}{dx}\displaystyle\int_{0}^{\sin x}\frac{dt}{x^5+t^5}$. (Your answer will still contain one unevaluated integral.)

2.4.16 Calculate $\dfrac{d}{dx}\displaystyle\int_{x}^{x^2} e^{-xt^2}\,dt$. (There will be one "impossible" integral left in your answer.)

2.4.17 Evaluate $\dfrac{d}{dx}\displaystyle\int_{0}^{x^2}\frac{\sin(xt)}{t}\,dt$. (Differentiate *first*, then do the surviving integral.)

2.4.18 Calculate

$$\frac{\partial}{\partial t}\int_{x-2t}^{x+2t}\sin(t+u)\,e^{-u^2}\,du,$$

using the chain rule and the fundamental theorem of calculus.

2.4.19 Find $\dfrac{d}{dt}\displaystyle\int_{t}^{t^2}(x^4-20)^{-1}\,dx$. (Don't evaluate any integrals.)

2.4.20 Give an alternative proof of the product rule (Example 1) by taking

$$\vec{F}(x) = \begin{pmatrix} f(x) \\ g(x) \end{pmatrix}, \qquad G(\vec{y}) = y_1 y_2\,.$$

2.5 Elementary Determinants

Associated with every square matrix is a number called its *determinant*. Indeed, one of the peculiarities of mathematics education is that many students are introduced to determinants several years before they encounter the matrices themselves. We are going to postpone a thorough study of the properties and significance of determinants until Chapter 7. However, in the meantime we will occasionally need determinants for incidental calculational purposes, so here we provide a quick review of how to calculate determinants in dimensions 2 and 3.

The determinant of a 2×2 matrix is the product of the elements on the main diagonal (the upper left and lower right), minus the product of the other two elements:

$$\begin{vmatrix} 2 & 3 \\ 1 & 4 \end{vmatrix} \equiv \det \begin{pmatrix} 2 & 3 \\ 1 & 4 \end{pmatrix} \equiv (2)(4) - (3)(1) = 5.$$

$$\begin{vmatrix} 2 & 3 \\ 4 & 6 \end{vmatrix} = (2)(6) - (3)(4) = 0.$$

The second of these examples demonstrates the principle that the determinant is zero if and only if the matrix is singular — that is, in dimension 2, if one row is a multiple of the other (see Exercise 2.5.8). The extension of this principle to higher dimensions will be very important to us in Chapters 4 and 5.

Often one needs to deal with determinants whose elements are functions or algebraic expressions. Here is a neat example:

$$\begin{vmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{vmatrix} = \cos^2\alpha + \sin^2\alpha = 1.$$

Two algorithms for evaluating 3×3 determinants are commonly taught. The first is the *cofactor expansion*, which reduces the calculation to the evaluation of three 2×2 determinants. For example, let us calculate the determinant of the matrix

$$A = \begin{pmatrix} 8 & 7 & 1 \\ 4 & 3 & 1 \\ -1 & -2 & 1 \end{pmatrix}.$$

$$\det A = \begin{vmatrix} 8 & 7 & 1 \\ 4 & 3 & 1 \\ -1 & -2 & 1 \end{vmatrix} = 8 \begin{vmatrix} 3 & 1 \\ -2 & 1 \end{vmatrix} - 4 \begin{vmatrix} 7 & 1 \\ -2 & 1 \end{vmatrix} + (-1) \begin{vmatrix} 7 & 1 \\ 3 & 1 \end{vmatrix}$$

$$= 8(3 \cdot 1 - (-2) \cdot 1) - 4(7 \cdot 1 - (-2) \cdot 1) - (7 \cdot 1 - 3 \cdot 1) = 40 - 36 - 4 = 0.$$

Here we have *expanded in cofactors of the first column*. Notice that the term corresponding to the *second* (middle) element of that column carries an extra minus sign. Exactly the same numerical result is obtained if one uses the first *row*, or, indeed, *any row or column* of the matrix as the foundation of the expansion, except that all the signs change if the *middle* row or column is chosen. (A perhaps less mysterious description of the sign rules in given in Sec. 7.1 in the context of a matrix of arbitrary size.) In practice, in hand calculation one chooses the basic row or column to be the one that promises the least messy arithmetic.

The second method is the 3-dimensional version of the *permutational definition* of the determinant. Here one writes down all 6 diagonal products of the matrix elements, attaching a minus sign to those that "slant upward":

$$\begin{vmatrix} 8 & 7 & 1 \\ 4 & 3 & 1 \\ -1 & -2 & 1 \end{vmatrix} = 8\cdot 3\cdot 1 + 7\cdot 1\cdot(-1) + 1\cdot 4\cdot(-2) - 1\cdot 3\cdot(-1) - 4\cdot 7\cdot 1 - 8\cdot(-2)\cdot 1$$

$$= 24 - 7 - 8 + 3 - 28 + 16 = 0.$$

This prescription is clearer if one thinks of the matrix as being written on a cylinder, which we can then unwrap so that we see the first two columns twice:

$$\left|\begin{array}{ccccc|cccc} 8 & & 7 & & 1 & & 8 & & 7 \\ & \searrow & & \mathrel{\nearrow\mkern-18mu\searrow} & & \mathrel{\nearrow\mkern-18mu\searrow} & & \nearrow & \\ 4 & & 3 & & 1 & & 4 & & 3 \\ & \nearrow & & \mathrel{\nearrow\mkern-18mu\searrow} & & \mathrel{\nearrow\mkern-18mu\searrow} & & \searrow & \\ -1 & & -2 & & 1 & & -1 & & -2 \end{array}\right|.$$

We close with three examples, each of which has a deeper significance in terms of applications of matrices that we will see later.

Example 1. Evaluate $\begin{vmatrix} \cos\alpha\sin\beta & \rho\cos\alpha\cos\beta & -\rho\sin\alpha\sin\beta \\ \sin\alpha\sin\beta & \rho\sin\alpha\cos\beta & \rho\cos\alpha\sin\beta \\ \cos\beta & -\rho\sin\beta & 0 \end{vmatrix}$.

SOLUTION: As a shortcut, we note that the factor $\rho\sin\beta$ appears in every element of the third column and therefore in every term of the answer; similarly, we can factor a ρ out of the second column. (See Sec. 7.1 for a formal statement of this property of determinants.) Now expand in cofactors of the bottom row:

$$\rho^2\sin\beta\begin{vmatrix} \cos\alpha\sin\beta & \cos\alpha\cos\beta & -\sin\alpha \\ \sin\alpha\sin\beta & \sin\alpha\cos\beta & \cos\alpha \\ \cos\beta & -\sin\beta & 0 \end{vmatrix}$$

$$= \rho^2\sin\beta[\cos\beta(\cos^2\alpha\cos\beta + \sin^2\alpha\cos\beta) + \sin\beta(\cos^2\alpha\sin\beta + \sin^2\alpha\sin\beta)]$$

$$= \rho^2\sin\beta[\cos^2\beta(\cos^2\alpha + \sin^2\alpha) + \sin^2\beta(\cos^2\alpha + \sin^2\alpha)]$$

$$= \rho^2\sin\beta(\cos^2\beta + \sin^2\beta) = \rho^2\sin\beta.$$

This determinant arises in calculating multiple integrals in spherical coordinates — see Sec. 7.3.

Example 2. Solve the equation $\begin{vmatrix} 1 & 1 & 1 \\ x & 2 & 3 \\ x^2 & 4 & 9 \end{vmatrix} = 0.$

SOLUTION: By cofactors of the first column,

$$\begin{vmatrix} 1 & 1 & 1 \\ x & 2 & 3 \\ x^2 & 4 & 9 \end{vmatrix} = 1\begin{vmatrix} 2 & 3 \\ 4 & 9 \end{vmatrix} - x\begin{vmatrix} 1 & 1 \\ 4 & 9 \end{vmatrix} + x^2\begin{vmatrix} 1 & 1 \\ 2 & 3 \end{vmatrix}$$

$$= (18 - 12) - x(9 - 4) + x^2(3 - 2) = x^2 - 5x + 6 = 0.$$

So the solutions are $x_1 = 2$ and $x_2 = 3$. Note that the solutions are closely related to the elements of the matrix itself; this is an example of a *Vandermonde matrix*, a structure that arises often in deriving formulas for numerical integration and interpolation (see several exercises in Sec. 7.1).

Example 3. Solve the equation $\det(A - \lambda) = 0$, where

$$A = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & -\frac{1}{2} & 0 \end{pmatrix}.$$

(Recall that $-\lambda$ here means $-\lambda I$, where I is the 3×3 identity matrix.)

SOLUTION: $(A - \lambda) = \begin{pmatrix} \frac{1}{4} - \lambda & \frac{1}{4} & -\frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} - \lambda & \frac{1}{4} \\ \frac{1}{2} & -\frac{1}{2} & -\lambda \end{pmatrix}.$

$$\det(A - \lambda) = \begin{vmatrix} \frac{1}{4} - \lambda & \frac{1}{4} & -\frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} - \lambda & \frac{1}{4} \\ \frac{1}{2} & -\frac{1}{2} & -\lambda \end{vmatrix} = \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{2} \begin{vmatrix} 1 - 4\lambda & 1 & -1 \\ 1 & 1 - 4\lambda & 1 \\ 1 & -1 & -2\lambda \end{vmatrix}$$

$$= \tfrac{1}{32}(-2\lambda(1 - 4\lambda)^2 + 1 + 1(-1)(-1) + (1 - 4\lambda) + (1 - 4\lambda) + 2\lambda)$$

$$= \tfrac{1}{32}(-2\lambda + 16\lambda^2 - 32\lambda^3 + 2 + 2 - 8\lambda + 2\lambda)$$

$$= \tfrac{1}{32}(-32\lambda^3 + 16\lambda^2 - 8\lambda + 4) = -\tfrac{1}{8}(2\lambda - 1)(4\lambda^2 + 1).$$

So $\lambda_1 = \frac{1}{2}$, $\lambda_2 = \frac{1}{2}i$, $\lambda_3 = -\frac{1}{2}i$, where $i^2 = -1$. These three roots are called the *eigenvalues* of A; see Chapter 8.

The cross product

Recall that in three-dimensional space there is a way of multiplying two vectors to get a third vector. (We delay to Chapter 7 an explanation of what is so special about dimension 3, and what happens to this *cross product* in other dimensions.) We already needed to state the definition of the cross product in Sec. 1.2, but here it is again, in alternative notation:

$$\begin{aligned}\vec{r}_1 \times \vec{r}_2 &= (y_1 z_2 - y_2 z_1)\hat{\imath} + (z_1 x_2 - z_2 x_1)\hat{\jmath} + (x_1 y_2 - x_2 y_1)\hat{k} \\ &= \begin{vmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{vmatrix}.\end{aligned}$$

The determinantal version of this formula requires some explanation, since the elements in the top row are vectors, not numbers, and a determinant of that nature has not been defined. (A function or formula involving numbers does not automatically make sense when the numbers are replaced by vectors. For example, what could $\vec{u}/\vec{v}$ mean, if $\vec{u}$ and $\vec{v}$ are not parallel?) However, if you expand the determinant by cofactors of the first row, interpreting the outermost multiplications in the obvious way as products of a vector with a scalar, you get the correct formula for the cross product. In the minds of most people, the determinant formula is easier to remember than the other formula with all the correct signs and coordinates.

Theorem: The cross product satisfies the identities (for all vectors $\vec{u}, \ldots \in \mathbf{R}^3$)

(1) $\vec{u} \times \vec{v} = -\,\vec{v} \times \vec{u}$;

(2) $\vec{u} \times (\vec{v} + \vec{w}) = \vec{u} \times \vec{v} + \vec{u} \times \vec{w}$, and the similar identity on the other side;

(3) $(r\vec{u}) \times \vec{v} = r(\vec{u} \times \vec{v}) = \vec{u} \times (r\vec{v})$ for any $r \in \mathbf{R}$;

(4) $\vec{u} \times (\vec{v} \times \vec{w}) + \vec{v} \times (\vec{w} \times \vec{u}) + \vec{w} \times (\vec{u} \times \vec{v}) = 0$;

(5) $\vec{u} \times (\vec{v} \times \vec{w}) = (\vec{u} \cdot \vec{w})\vec{v} - (\vec{u} \cdot \vec{v})\vec{w}$ and $(\vec{u} \times \vec{v}) \times \vec{w} = (\vec{u} \cdot \vec{w})\vec{v} - (\vec{v} \cdot \vec{w})\vec{u}$;

(6) $\vec{u} \cdot (\vec{v} \times \vec{w}) = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$, the determinant of the matrix whose rows (or columns) are the vectors concerned.

Note from properties (1) and (5) that the cross product is neither commutative nor associative:

$$\vec{u} \times \vec{v} \neq \vec{v} \times \vec{u}, \qquad \vec{u} \times (\vec{v} \times \vec{w}) \neq (\vec{u} \times \vec{v}) \times \vec{w} = -\,\vec{w} \times (\vec{u} \times \vec{v}).$$

The structure of both identities in (5) is easy to remember from the slogan

outside dot remote times adjacent,

minus outside dot adjacent times remote,

where "adjacent" and "remote" describe the positions of the vectors inside the parentheses relative to the outside vector. The combinations of three vectors appearing in (4) and (5) are called *vector triple products*, and those of type (6) are called *scalar triple products*.

Exercises

2.5.1 Calculate the determinants of these matrices:

$$A = \begin{pmatrix} 2 & 5 \\ 3 & 8 \end{pmatrix}, \qquad B = \begin{pmatrix} -1 & -3 \\ 2 & 4 \end{pmatrix},$$

$$C = \begin{pmatrix} 2 & 5 \\ -3 & 4 \end{pmatrix}, \qquad D = \begin{pmatrix} -7 & 12 \\ -3 & 4 \end{pmatrix}.$$

2.5.2 Calculate the determinants of these matrices:

$$A = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 1 & 2 \\ 3 & 4 & 5 \end{pmatrix}, \qquad B = \begin{pmatrix} 2 & 1 & 3 \\ -3 & 2 & -1 \\ 5 & -3 & -2 \end{pmatrix}.$$

2.5.3 Calculate the determinants of these matrices:

$$C = \begin{pmatrix} 1 & 7 & 5 \\ -3 & 2 & -1 \\ 5 & -3 & -2 \end{pmatrix}, \qquad D = \begin{pmatrix} 2 & 1 & 3 \\ 1 & 7 & 5 \\ 5 & -3 & -2 \end{pmatrix}.$$

2.5.4 Calculate the determinants of these matrices:

$$A = \begin{pmatrix} \alpha+\beta & \alpha-\beta \\ \alpha-\beta & \alpha+\beta \end{pmatrix}, \qquad B = \begin{pmatrix} x^2-x+1 & x \\ x^2 & x+1 \end{pmatrix}.$$

2.5.5 Calculate the determinant of $\begin{pmatrix} \frac{\cos^2\alpha}{1+\sin^2\alpha} & \frac{-2\sin\alpha}{1+\sin^2\alpha} \\ \frac{2\sin\alpha}{1+\sin^2\alpha} & \frac{\cos^2\alpha}{1+\sin^2\alpha} \end{pmatrix}$.

2.5.6 Calculate the determinant of $\begin{vmatrix} x_1 & 2 & 3 \\ x_2 & -2 & 3 \\ x_3 & 1 & -1 \end{vmatrix}$.

2.5.7 Solve the equation $\begin{vmatrix} 1 & 2 & 8-x \\ 1 & 5-x & 3 \\ 1 & 2 & 3 \end{vmatrix} = 0.$

2.5.8 Prove that the determinant of a 2×2 matrix is 0 if and only if one row of the matrix is proportional to the other row. What can you say about proportionality of the *columns*?

2.5.9 Let $\vec{u} = \hat{\imath} - \hat{\jmath} + 2\hat{k}, \;\; \vec{v} = 5\hat{\imath} + 2\hat{\jmath} + \hat{k}$.

(a) Calculate $\vec{u} \times \vec{v}$.

(b) Verify that $\|\vec{u} \times \vec{v}\| = \sqrt{\|\vec{u}\|^2\|\vec{v}\|^2 - (\vec{u} \cdot \vec{v})^2}$.

2.5.10

(a) For arbitrary vectors in $\mathbf{R}^3$, show by algebraic calculation that

$$\|\vec{u} \times \vec{v}\|^2 = \|\vec{u}\|^2\|\vec{v}\|^2 - (\vec{u} \cdot \vec{v})^2.$$

(b) Use the identity in (a) to show that

$$\|\vec{u} \times \vec{v}\| = \|\vec{u}\|\|\vec{v}\| \sin\theta,$$

where θ is the angle between $\vec{u}$ and $\vec{v}$.

2.5.11 Is it possible to define a vector division operation inverse to the cross product, so that

$$\frac{\vec{u} \times \vec{v}}{\vec{v}} = \vec{u}$$

for all $\vec{u}$ and nonzero $\vec{v}$ in $\mathbf{R}^3$?

2.5.12 Prove one of the "outside dot remote … " identities (5), and deduce the other one from it.